

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant(s): MORIMOTO, Hiroyuki  
Serial No.: Not yet assigned  
Filed: April 5, 2004  
Title: STORAGE UNIT, STORAGE UNIT CONTROL METHOD, AND  
STORAGE SYSTEM  
Group: Not yet assigned

LETTER CLAIMING RIGHT OF PRIORITY

Commissioner for Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

April 5, 2004

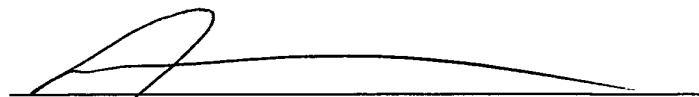
Sir:

Under the provisions of 35 USC 119 and 37 CFR 1.55, the applicant(s) hereby claim(s) the right of priority based on Japanese Patent Application No.(s) 2004-019739, filed January 28, 2004.

A certified copy of said Japanese Application is attached.

Respectfully submitted,

ANTONELLI, TERRY, STOUT & KRAUS, LLP



Alan E. Schiavelli  
Registration No. 32,087

AES/alb  
Attachment  
(703) 312-6600

日 本 国 特 許 庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日                      2 0 0 4 年    1 月 2 8 日  
Date of Application:

出 願 番 号                      特 願 2 0 0 4 - 0 1 9 7 3 9  
Application Number:

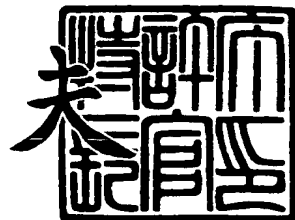
[ST. 10/C]:                      [ J P 2 0 0 4 - 0 1 9 7 3 9 ]

出      願      人                      株式会社日立製作所  
Applicant(s):

2 0 0 4 年    3 月 1 8 日

特許庁長官  
Commissioner,  
Japan Patent Office

今 井 康 夫



出証番号    出証特 2 0 0 4 - 3 0 2 2 0 8 3

【書類名】 特許願  
【整理番号】 340301445  
【提出日】 平成16年 1月28日  
【あて先】 特許庁長官殿  
【国際特許分類】 G06F 3/06  
【発明者】  
    【住所又は居所】 神奈川県小田原市中里 3 2 2 番 2 号 株式会社日立製作所 R A I  
                            Dシステム事業部内  
    【氏名】 森本 浩之  
【特許出願人】  
    【識別番号】 000005108  
    【氏名又は名称】 株式会社日立製作所  
【代理人】  
    【識別番号】 110000176  
    【氏名又は名称】 一色国際特許業務法人  
    【代表者】 一色 健輔  
【手数料の表示】  
    【予納台帳番号】 211868  
    【納付金額】 21,000円  
【提出物件の目録】  
    【物件名】 特許請求の範囲 1  
    【物件名】 明細書 1  
    【物件名】 図面 1  
    【物件名】 要約書 1

**【書類名】 特許請求の範囲****【請求項 1】**

データを記憶する複数の第 1 のハードディスクドライブを有する複数の他のストレージ装置と通信可能に接続され、

データを記憶する複数の第 2 のハードディスクドライブと、

前記他のストレージ装置のそれぞれから、前記複数の第 1 のハードディスクドライブのデータ記憶領域を論理的に区分してなる複数の記憶ブロックのそれぞれに記憶される第 1 の記憶データの複製を、前記各記憶ブロックを特定する第 1 の識別子と共に受信する第 1 の受信部と、

前記第 1 の受信部により前記他のストレージ装置のそれぞれから受信された前記第 1 の記憶データの複製のうち、前記第 1 の識別子がそれぞれ対応する前記第 1 の記憶データの複製の排他的論理和を演算する第 1 の演算制御部と、

前記複数の第 2 のハードディスクドライブのデータ記憶領域を論理的に区分してなる複数の記憶ブロックのそれぞれを特定する第 2 の識別子が前記第 1 の識別子と対応する前記第 2 のハードディスクドライブの前記記憶ブロックに、前記第 1 の演算制御部により演算された前記排他的論理和の演算結果を記憶する第 1 の記憶制御部とを備えることを特徴とするストレージ装置。

**【請求項 2】**

前記第 2 のハードディスクドライブには、前記第 1 の演算制御部により演算された前記排他的論理和の全ての演算結果が記憶され、

前記第 1 のハードディスクドライブへの書き込みデータを情報処理装置から受信した前記他のストレージ装置により演算される、前記書き込みデータと前記書き込みデータが書き込まれる前記第 1 のハードディスクドライブの前記記憶ブロックに記憶されている前記第 1 の記憶データとの排他的論理和の演算結果を、前記書き込みデータが書き込まれる前記記憶ブロックを特定する前記第 1 の識別子と共に、前記他のストレージ装置から受信する第 2 の受信部と、

前記第 2 の受信部により受信された前記演算結果と、前記第 2 の受信部により受信された前記第 1 の識別子と対応する前記第 2 の識別子で特定される前記第 2 のハードディスクドライブの前記記憶ブロックに記憶されている第 2 の記憶データとの排他的論理和を演算する第 2 の演算制御部と、

前記第 2 の演算制御部により演算された前記排他的論理和の演算結果を、前記第 2 の識別子で特定される前記第 2 のハードディスクドライブの前記記憶ブロックに記憶する第 2 の記憶制御部と

を備えることを特徴とする請求項 1 に記載のストレージ装置。

**【請求項 3】**

前記第 2 のハードディスクドライブには、前記第 1 の演算制御部により演算された前記排他的論理和の全ての演算結果が記憶され、

通信可能に接続される前記他のストレージ装置が追加された場合に、

前記追加された他のストレージ装置から、前記追加された他のストレージ装置が有する前記第 1 のハードディスクドライブの前記記憶ブロックのそれぞれに記憶される前記第 1 の記憶データの複製を、前記第 1 の識別子と共に受信する第 3 の受信部と、

前記第 3 の受信部により受信された前記第 1 の記憶データの複製と、前記第 3 の受信部により受信された前記第 1 の識別子と対応する前記第 2 の識別子で特定される前記第 2 のハードディスクドライブの前記記憶ブロックに記憶されている第 2 の記憶データとの排他的論理和を演算する第 3 の演算制御部と、

前記第 3 の演算制御部により演算された前記排他的論理和の演算結果を、前記第 2 の識別子で特定される前記第 2 のハードディスクドライブの前記記憶ブロックに記憶する第 3 の記憶制御部と

を備えることを特徴とする請求項 1 に記載のストレージ装置。

**【請求項 4】**

前記第2のハードディスクドライブには、前記第1の演算制御部により演算された前記排他的論理和の全ての演算結果が記憶され、

通信可能に接続される前記他のストレージ装置が追加された場合に、

前記追加された他のストレージ装置が有する前記第1のハードディスクドライブへの書き込みデータを情報処理装置から受信した前記追加された他のストレージ装置により演算される、前記書き込みデータと前記書き込みデータが書き込まれる前記第1のハードディスクドライブの前記記憶ブロックに記憶されている前記第1の記憶データとの排他的論理和の演算結果を、前記書き込みデータが書き込まれる前記記憶ブロックを特定する前記第1の識別子と共に、前記追加された他のストレージ装置から受信する第4の受信部と、

前記第4の受信部により受信された前記演算結果と、前記第4の受信部により受信された前記第1の識別子と対応する前記第2の識別子で特定される前記第2のハードディスクドライブの前記記憶ブロックに記憶されている第2の記憶データとの排他的論理和を演算する第4の演算制御部と、

前記第4の演算制御部により演算された前記排他的論理和の演算結果を、前記第2の識別子で特定される前記第2のハードディスクドライブの前記記憶ブロックに記憶する第4の記憶制御部と

を備えることを特徴とする請求項1に記載のストレージ装置。

#### 【請求項5】

前記第2のハードディスクドライブには、前記第1の演算制御部により演算された前記排他的論理和の全ての演算結果が記憶され、

ある前記他のストレージ装置から、前記ある他のストレージ装置が有する前記第1のハードディスクドライブに記憶されているべき前記第1の記憶データの送信要求を受信する第5の受信部と、

前記ある他のストレージ装置以外の前記他のストレージ装置に、それぞれの前記第1のハードディスクドライブの前記各記憶ブロックに記憶されている前記第1の記憶データの複製を、前記各第1の記憶データが記憶されている前記記憶ブロックを特定する前記第1の識別子と共に送信させるための送信要求を送信する第1の送信部と、

前記ある他のストレージ装置以外の前記他のストレージ装置のそれぞれから、前記第1の記憶データの複製を、前記第1の識別子と共に受信する第6の受信部と、

前記第6の受信部により前記ある他のストレージ装置以外の前記他のストレージ装置のそれぞれから受信された前記第1の記憶データの複製のうち、前記第1の識別子がそれぞれ対応する前記第1の記憶データの複製と、前記第6の受信部により受信された前記第1の識別子と対応する前記第2の識別子で特定される前記第2のハードディスクドライブの前記記憶ブロックに記憶されている第2の記憶データとの排他的論理和を演算する第5の演算制御部と、

前記第5の演算制御部により演算された前記排他的論理和の演算結果を、前記第1の識別子と共に、前記ある他のストレージ装置に送信する第2の送信部とを備えることを特徴とする請求項1に記載のストレージ装置。

#### 【請求項6】

データを記憶する複数の第1のハードディスクドライブを有する複数の他のストレージ装置と通信可能に接続され、データを記憶する複数の第2のハードディスクドライブを備えるストレージ装置の制御方法であって、

前記他のストレージ装置のそれぞれから、前記複数の第1のハードディスクドライブのデータ記憶領域を論理的に区分してなる複数の記憶ブロックのそれぞれに記憶される第1の記憶データの複製を、前記各記憶ブロックを特定する第1の識別子と共に受信し、

前記他のストレージ装置のそれぞれから受信した前記第1の記憶データの複製のうち、前記第1の識別子がそれぞれ対応する前記第1の記憶データの複製の排他的論理和を演算し、

前記複数の第2のハードディスクドライブのデータ記憶領域を論理的に区分してなる複数の記憶ブロックのそれぞれを特定する第2の識別子が前記第1の識別子と対応する前記

第2のハードディスクドライブの前記記憶ブロックに、前記排他的論理和の演算結果を記憶すること

を特徴とするストレージ装置の制御方法。

【請求項7】

前記第2のハードディスクドライブには、前記排他的論理和の全ての演算結果が記憶され、

前記第1のハードディスクドライブへの書き込みデータを情報処理装置から受信した前記他のストレージ装置により演算される、前記書き込みデータと前記書き込みデータが書き込まれる前記第1のハードディスクドライブの前記記憶ブロックに記憶されている前記第1の記憶データとの排他的論理和の演算結果を、前記書き込みデータが書き込まれる前記記憶ブロックを特定する前記第1の識別子と共に、前記他のストレージ装置から受信し、

前記演算結果と、前記第1の識別子と対応する前記第2の識別子で特定される前記第2のハードディスクドライブの前記記憶ブロックに記憶されている第2の記憶データとの排他的論理和を演算し、

前記排他的論理和の演算結果を、前記第2の識別子で特定される前記第2のハードディスクドライブの前記記憶ブロックに記憶すること

を特徴とする請求項6に記載のストレージ装置の制御方法。

【請求項8】

前記第2のハードディスクドライブには、前記排他的論理和の全ての演算結果が記憶され、

通信可能に接続される前記他のストレージ装置が追加された場合に、

前記追加された他のストレージ装置から、前記追加された他のストレージ装置が有する前記第1のハードディスクドライブの前記記憶ブロックのそれぞれに記憶される前記第1の記憶データの複製を、前記第1の識別子と共に受信し、

前記第1の記憶データの複製と、前記第1の識別子と対応する前記第2の識別子で特定される前記第2のハードディスクドライブの前記記憶ブロックに記憶されている第2の記憶データとの排他的論理和を演算し、

前記排他的論理和の演算結果を、前記第2の識別子で特定される前記第2のハードディスクドライブの前記記憶ブロックに記憶すること

を特徴とする請求項6に記載のストレージ装置の制御方法。

【請求項9】

前記第2のハードディスクドライブには、前記排他的論理和の全ての演算結果が記憶され、

通信可能に接続される前記他のストレージ装置が追加された場合に、

前記追加された他のストレージ装置が有する前記第1のハードディスクドライブへの書き込みデータを情報処理装置から受信した前記追加された他のストレージ装置により演算される、前記書き込みデータと前記書き込みデータが書き込まれる前記第1のハードディスクドライブの前記記憶ブロックに記憶されている前記第1の記憶データとの排他的論理和の演算結果を、前記書き込みデータが書き込まれる前記記憶ブロックを特定する前記第1の識別子と共に、前記追加された他のストレージ装置から受信し、

前記演算結果と、前記第1の識別子と対応する前記第2の識別子で特定される前記第2のハードディスクドライブの前記記憶ブロックに記憶されている第2の記憶データとの排他的論理和を演算し、

前記排他的論理和の演算結果を、前記第2の識別子で特定される前記第2のハードディスクドライブの前記記憶ブロックに記憶すること

を特徴とする請求項6に記載のストレージ装置の制御方法。

【請求項10】

前記第2のハードディスクドライブには、前記排他的論理和の全ての演算結果が記憶され、

ある前記他のストレージ装置から、前記ある他のストレージ装置が有する前記第1のハードディスクドライブに記憶されているべき前記第1の記憶データの送信要求を受信し、

前記ある他のストレージ装置以外の前記他のストレージ装置に、それぞれの前記第1のハードディスクドライブの前記各記憶ブロックに記憶されている前記第1の記憶データの複製を、前記各第1の記憶データが記憶されている前記記憶ブロックを特定する前記第1の識別子と共に送信させるための送信要求を送信し、

前記ある他のストレージ装置以外の前記他のストレージ装置のそれぞれから、前記第1の記憶データの複製を、前記第1の識別子と共に受信し、

前記ある他のストレージ装置以外の前記他のストレージ装置のそれぞれから受信した前記第1の記憶データの複製のうち、前記第1の識別子がそれぞれ対応する前記第1の記憶データの複製と、前記第1の識別子と対応する前記第2の識別子で特定される前記第2のハードディスクドライブの前記記憶ブロックに記憶されている第2の記憶データとの排他的論理和を演算し、

前記排他的論理和の演算結果を、前記第1の識別子と共に、前記ある他のストレージ装置に送信すること

を特徴とする請求項6に記載のストレージ装置の制御方法。

#### 【請求項11】

データを記憶する複数の第1のハードディスクドライブを有する複数の第1のストレージ装置と、前記各第1のストレージ装置と通信可能に接続され、データを記憶する複数の第2のハードディスクドライブを有する第2のストレージ装置とを備え、

前記第1のストレージ装置は、

前記複数の第1のハードディスクドライブのデータ記憶領域を論理的に区分してなる複数の記憶ブロックのそれぞれに記憶される第1の記憶データの複製を、前記各記憶ブロックを特定する第1の識別子と共に前記第2のストレージ装置に送信する第1のデータ送信部を備え、

前記第2のストレージ装置は、

前記第1のストレージ装置のそれぞれから、前記第1の記憶データの複製を、前記第1の識別子と共に受信する第1のデータ受信部と、

前記第1のデータ受信部により前記第1のストレージ装置のそれぞれから受信された前記第1の記憶データの複製のうち、前記第1の識別子がそれぞれ対応する前記第1の記憶データの複製の排他的論理和を演算する第1のデータ演算制御部と、

前記複数の第2のハードディスクドライブのデータ記憶領域を論理的に区分してなる複数の記憶ブロックのそれぞれを特定する第2の識別子が前記第1の識別子と対応する前記第2のハードディスクドライブの前記記憶ブロックに、前記第1のデータ演算制御部により演算された前記排他的論理和の演算結果を記憶する第1のデータ記憶制御部とを備えること

を特徴とするストレージシステム。

#### 【請求項12】

前記第2のハードディスクドライブには、前記第1のデータ演算制御部により演算された前記排他的論理和の全ての演算結果が記憶され、

前記第1のストレージ装置は、

情報処理装置から前記第1のハードディスクドライブへの書き込みデータを受信した場合に、前記書き込みデータと前記書き込みデータが書き込まれる前記第1のハードディスクドライブの前記記憶ブロックに記憶されている前記第1の記憶データとの排他的論理和を演算する第2のデータ演算制御部と、

前記第2のデータ演算制御部により演算された前記排他的論理和の演算結果を、前記書き込みデータが書き込まれる前記記憶ブロックを特定する前記第1の識別子と共に、前記第2のストレージ装置に送信する第2のデータ送信部と、

を備え、

前記第2のストレージ装置は、

前記第2のデータ演算制御部により演算された前記排他的論理和の演算結果を、前記第1のストレージ装置から、前記第1の識別子と共に受信する第2のデータ受信部と、

前記第2のデータ受信部により受信された前記演算結果と、前記第2のデータ受信部により受信された前記第1の識別子と対応する前記第2の識別子で特定される前記第2のハードディスクドライブの前記記憶ブロックに記憶されている第2の記憶データとの排他的論理和を演算する第3のデータ演算制御部と、

前記第3のデータ演算制御部により演算された前記排他的論理和の演算結果を、前記第2の識別子で特定される前記第2のハードディスクドライブの前記記憶ブロックに記憶する第2のデータ記憶制御部と

を備えること

を特徴とする請求項11に記載のストレージシステム。

【請求項13】

前記第2のハードディスクドライブには、前記第1のデータ演算制御部により演算された前記排他的論理和の全ての演算結果が記憶され、

前記第2のストレージ装置は、

前記第2のストレージ装置と通信可能に接続される前記第1のストレージ装置が追加された場合に、前記追加された第1のストレージ装置に、前記第1のハードディスクドライブの前記各記憶ブロックに記憶されている前記第1の記憶データの複製を、前記各第1の記憶データが記憶されている前記記憶ブロックを特定する前記第1の識別子と共に送信させるための送信要求を送信する第3のデータ送信部を備え、

前記第1のストレージ装置は、

前記送信要求に応じて前記第1のハードディスクドライブの前記各記憶ブロックに記憶されている前記第1の記憶データの複製を、前記第1の識別子と共に前記第2のストレージ装置に送信する第4のデータ送信部を備え、

前記第2のストレージ装置は、

前記第1の記憶データの複製を、前記第1の識別子と共に前記第1のストレージ装置から受信する第3のデータ受信部と、

前記第3のデータ受信部により受信された前記第1の記憶データの複製と、前記第3のデータ受信部により受信された前記第1の識別子と対応する前記第2の識別子で特定される前記第2のハードディスクドライブの前記記憶ブロックに記憶されている第2の記憶データとの排他的論理和を演算する第4のデータ演算制御部と、

前記第4のデータ演算制御部により演算された前記排他的論理和の演算結果を、前記第2の識別子で特定される前記第2のハードディスクドライブの前記記憶ブロックに記憶する第3のデータ記憶制御部と

を備えることを特徴とする請求項11に記載のストレージシステム。

【請求項14】

前記第2のハードディスクドライブには、前記第1のデータ演算制御部により演算された前記排他的論理和の全ての演算結果が記憶され、

前記第2のストレージ装置と通信可能に接続される前記第1のストレージ装置が追加された場合に、

前記追加された第1のストレージ装置は、

情報処理装置から前記第1のハードディスクドライブへの書き込みデータを受信した場合に、前記書き込みデータと前記書き込みデータが書き込まれる前記第1のハードディスクドライブの前記記憶ブロックに記憶されている前記第1の記憶データとの排他的論理和を演算する第5のデータ演算制御部と、

前記第5のデータ演算制御部により演算された前記排他的論理和の演算結果を、前記書き込みデータが書き込まれる前記記憶ブロックを特定する前記第1の識別子と共に、前記第2のストレージ装置に送信する第5のデータ送信部と、

を備え、

前記第2のストレージ装置は、



前記追加された第1のストレージ装置から、前記第5のデータ演算制御部により演算された前記排他的論理和の演算結果を前記第1の識別子と共に受信する第4のデータ受信部と、

前記第4のデータ受信部により受信された前記演算結果と、前記第4のデータ受信部により受信された前記第1の識別子と対応する前記第2の識別子で特定される前記第2のハードディスクドライブの前記記憶ブロックに記憶されている第2の記憶データとの排他的論理和を演算する第6のデータ演算制御部と、

前記第6のデータ演算制御部により演算された前記排他的論理和の演算結果を、前記第2の識別子で特定される前記第2のハードディスクドライブの前記記憶ブロックに記憶する第4のデータ記憶制御部と

を備えること

を特徴とする請求項11に記載のストレージシステム。

【請求項15】

前記第2のハードディスクドライブには、前記第1のデータ演算制御部により演算された前記排他的論理和の全ての演算結果が記憶され、

前記第1のストレージ装置は、

前記第1のストレージ装置が有する前記第1のハードディスクドライブに記憶されているべき前記第1の記憶データの送信要求を前記第2のストレージ装置に送信する第6のデータ送信部を備え、

前記第2のストレージ装置は、

前記第1のストレージ装置から、前記第1の記憶データの送信要求を受信する第5のデータ受信部と、

前記第1の記憶データの送信要求に応じて、前記第1の記憶データの送信要求を送信した前記第1のストレージ装置以外の前記第1のストレージ装置に、それぞれの前記第1のハードディスクドライブの前記各記憶ブロックに記憶されている前記第1の記憶データの複製を、前記各第1の記憶データが記憶されている前記記憶ブロックを特定する前記第1の識別子と共に送信させるための送信要求を送信する第7のデータ送信部と、

前記第1の記憶データの送信要求を送信した前記第1のストレージ装置以外の前記第1のストレージ装置のそれぞれから、前記第1の記憶データの複製を、前記第1の識別子と共に受信する第6のデータ受信部と、

前記第6のデータ受信部により受信された前記第1の記憶データの複製のうち、前記第1の識別子がそれぞれ対応する前記第1の記憶データの複製と、前記第6のデータ受信部により受信された前記第1の識別子と対応する前記第2の識別子で特定される前記第2のハードディスクドライブの前記記憶ブロックに記憶されている第2の記憶データとの排他的論理和を演算する第7のデータ演算制御部と、

前記第7のデータ演算制御部により演算された前記排他的論理和の演算結果を、前記第1の識別子と共に、前記第1の記憶データの送信要求を送信した前記第1のストレージ装置に送信する第8のデータ送信部と、

を備え、

前記第1のストレージ装置は、

前記第7のデータ演算制御部により演算された前記排他的論理和の演算結果を、前記第1の識別子と共に、前記第2のストレージ装置から受信する第7のデータ受信部と、

前記第7のデータ受信部により受信された前記排他的論理和の演算結果を、前記第7のデータ受信部により受信された前記第1の識別子で特定される前記第1のハードディスクドライブの前記記憶ブロックに記憶する第5のデータ記憶制御部と

を備えること

を特徴とする請求項11に記載のストレージシステム。

**【書類名】 明細書****【発明の名称】** ストレージ装置、ストレージ装置の制御方法、及びストレージシステム**【技術分野】****【0001】**

本発明は、ストレージ装置、ストレージ装置の制御方法、及びストレージシステムに関する。

**【背景技術】****【0002】**

ストレージ装置に記憶されるデータが消失した場合に、元のデータを復元できるようにするために、データのバックアップが行われる。このデータのバックアップは、通常ストレージ装置に記憶されるデータの複製を、バックアップ用のストレージ装置にも記憶させておくことにより行われる。

**【0003】**

またこのデータのバックアップをレプリケーションの技術を利用して行う技術も開発されている。レプリケーションは、通信可能に接続されたストレージ装置間で、一方のストレージ装置にデータが書き込まれると、そのデータの複製が他方のストレージ装置に送信され、他方のストレージ装置にそのデータの複製が書き込まれるように制御するための技術である。

**【特許文献1】** 特開 2002-259183 号公報

**【発明の開示】****【発明が解決しようとする課題】****【0004】**

しかしながらバックアップされる元のデータを記憶するストレージ装置が複数ある場合には、バックアップ用のストレージ装置は、それらのストレージ装置のそれぞれに記憶される元のデータの全ての複製を記憶することができるだけの記憶容量を備えていることが必要である。

またレプリケーションの技術を利用してデータのバックアップを行う場合には、バックアップされる元のデータを記憶するストレージ装置毎に、ペアを組むバックアップ用のストレージ装置を設ける必要がある。

このようなことから、データのバックアップを行うためには、数多くのストレージ装置を維持管理することが必要となり、そのための維持管理コストも大きなものとなる。

**【0005】**

今日、情報処理システムで取り扱われるデータ量は急激に増大しており、データのバックアップを合理化し、バックアップのためのコストを抑制し、ひいては情報処理システムのコストを抑制するための技術が強く求められている。

**【0006】**

本発明は上記課題を鑑みてなされたものであり、ストレージ装置、ストレージ装置の制御方法、及びストレージシステムを提供することを主たる目的とする。

**【課題を解決するための手段】****【0007】**

上記課題を解決するために、本発明は、データを記憶する複数の第1のハードディスクドライブを有する複数の他のストレージ装置と通信可能に接続され、データを記憶する複数の第2のハードディスクドライブと、前記他のストレージ装置のそれぞれから、前記複数の第1のハードディスクドライブのデータ記憶領域を論理的に区分してなる複数の記憶ブロックのそれぞれに記憶される第1の記憶データの複製を、前記各記憶ブロックを特定する第1の識別子と共に受信する第1の受信部と、前記第1の受信部により前記他のストレージ装置のそれぞれから受信された前記第1の記憶データの複製のうち、前記第1の識別子がそれぞれ対応する前記第1の記憶データの複製の排他的論理和を演算する第1の演算制御部と、前記複数の第2のハードディスクドライブのデータ記憶領域を論理的に区分してなる複数の記憶ブロックのそれぞれを特定する第2の識別子が前記第1の識別子と対

応する前記第2のハードディスクドライブの前記記憶ブロックに、前記第1の演算制御部により演算された前記排他的論理和の演算結果を記憶する第1の記憶制御部とを備えることを特徴とするストレージ装置に関する。

【0008】

その他、本願が開示する課題、及びその解決方法は、発明を実施するための最良の形態の欄、及び図面により明らかにされる。

【発明の効果】

【0009】

ストレージ装置、ストレージ装置の制御方法、及びストレージシステムを提供することができる。

【発明を実施するための最良の形態】

【0010】

===全体構成例===

まず、本実施の形態に係るストレージ装置600を含む情報処理システムの全体構成を示すブロック図を図1に示す。本実施の形態に係る情報処理システムは、複数のストレージ装置600を含んで構成されるストレージシステムと、各ストレージ装置600と通信可能に接続される情報処理装置200とを含んで構成される。

【0011】

本実施の形態に係るストレージシステムは、ストレージ装置1乃至4(600)を備える。ストレージ装置1乃至3(他のストレージ装置、第1のストレージ装置、以下データストレージ装置とも記す)(600)には、情報処理装置1乃至4(200)が通信可能に接続されている。ストレージ装置1(600)には情報処理装置1(200)が接続され、ストレージ装置2(600)には情報処理装置2(200)が接続され、ストレージ装置3(600)には情報処理装置3及び4(200)が接続されている。各情報処理装置200は、それぞれ通信可能に接続されたストレージ装置600に対してデータ入出力要求を送信する。各ストレージ装置600は、情報処理装置200から送信されたデータ入出力要求に応じて、ストレージ装置600が備える記憶領域(データ記憶領域とも記す)に記憶されるデータの読み書きを行う。

【0012】

情報処理装置200はCPU(Central Processing Unit)やメモリを備えたコンピュータ等の情報機器である。情報処理装置200が備えるCPUにより各種アプリケーションプログラムが実行されることにより様々な機能が実現される。情報処理装置200により実現される機能としては、例えば銀行の自動預金預け払いサービスや、航空機の座席予約サービスとすることができる。情報処理装置200は、例えばパーソナルコンピュータやワークステーションとすることもできるし、メインフレームコンピュータとすることもできる。

【0013】

ストレージ装置3(600)は、SAN(Storage Area Network)500を介して情報処理装置3及び4(200)と通信可能に接続されている。SAN500を介して行われる情報処理装置200とストレージ装置600との間の通信は、ファイバチャネルプロトコルに従って行われるようにすることができる。この場合SAN500は、ファイバチャネルプロトコルに準拠した少なくとも一つ以上のスイッチ等の通信機器により構成される。情報処理装置3及び4(200)からは、ストレージ装置3(600)に対して、ファイバチャネルプロトコルに従ってデータ入出力要求が送信される。

【0014】

もちろん、SAN500によらずに情報処理装置3及び4(200)とストレージ装置3(600)とを接続するようにすることもできる。この場合、情報処理装置3及び4(200)とストレージ装置3(600)との間の通信は、例えばFICON(Fibre Connection)(登録商標)やESCON(Enterprise System Connection)(登録商標)、TCP/IP(Transmission Control Protocol / Internet Protocol)、iSCSI(Int

ernet Small Computer Systems Interface) などの通信プロトコルにより行うようにすることもできる。

#### 【0015】

また、もちろん情報処理装置 1 (200) とストレージ装置 1 (600) との間の通信、情報処理装置 2 (200) とストレージ装置 2 (600) との間の通信についても、ファイバチャネルプロトコルや、FICON (登録商標)、ESCON (登録商標)、TCP/IP、iSCSI などの通信プロトコルにより行うようにすることもできる。

#### 【0016】

また図 1 に示すように、情報処理装置 1 乃至 4 (200) は LAN (Local Area Network) 400 で通信可能に接続されているようにすることもできる。ここで LAN 400 は、例えばインターネットとすることができる。

#### 【0017】

また、各ストレージ装置 600 に接続される情報処理装置 200 の数は図 1 に示される例に限定されることはなく、適宜の台数にすることができる。また同様に、各情報処理装置 200 に接続されるストレージ装置 600 の数も、適宜の台数とすることができる。

#### 【0018】

ストレージ装置 4 (第 2 のストレージ装置、以下、パリティストレージ装置とも記す) (600) は、ストレージ装置 1 乃至 3 (600) と通信可能に接続されている。そして本実施の形態に係るストレージ装置 4 (600) が備える記憶領域には、ストレージ装置 1 乃至 3 (600) (他のストレージ装置) の記憶領域にそれぞれ記憶されるデータをバックアップするためのデータが記憶される。

#### 【0019】

すなわち、詳細な仕組みについては後述するが、本実施の形態に係るストレージ装置 4 (600) には、ストレージ装置 1 乃至 3 (600) の記憶領域にそれぞれ記憶されるデータから演算される排他的論理和の演算結果が記憶される。そしてストレージ装置 1 乃至 3 のいずれかの記憶領域に記憶されるデータ、例えばストレージ装置 1 (600) の記憶領域に記憶されるデータを復元することが必要となった場合には、ストレージ装置 1 (600) 以外のストレージ装置 (600)、すなわちストレージ装置 2 乃至 4 (600) の記憶領域に記憶されるデータから排他的論理和を演算することによりストレージ装置 1 (600) の記憶領域に記憶されるデータを復元するようにする。

#### 【0020】

これを式で表せば以下のようなになる。すなわち、ストレージ装置 1 (600) に記憶されるデータを D1、ストレージ装置 2 (600) に記憶されるデータを D2、ストレージ装置 3 (600) に記憶されるデータを D3 とすれば、ストレージ装置 4 (600) に記憶されるデータである D4 は、 $D1 (EXOR) D2 (EXOR) D3$  となる。ただし“(EXOR)”は排他的論理和の演算子を示す。ここで、D1 を復元することが必要となった場合には、 $D4 (EXOR) D2 (EXOR) D3$  を演算する。このようにして D1 を復元することができる。

#### 【0021】

このようにすることにより、本実施の形態によれば、ストレージ装置 4 (600) と通信可能に接続される他のストレージ装置 600 の台数が何台であっても、他のストレージ装置 (600) の記憶領域に記憶されるデータをバックアップすることが可能となる。

#### 【0022】

なおストレージ装置 4 (600) は、ストレージ装置 1 乃至 3 (600) の記憶領域にそれぞれ記憶されるデータをバックアップするためのデータを記憶するための専用のストレージ装置 600 とすることもできるし、ストレージ装置 1 乃至 3 (600) と同様に、情報処理装置 200 と通信可能に接続され、その情報処理装置 200 からのデータ入出力要求に応じて記憶領域に記憶されるデータの読み書きを行うこともできるストレージ装置 600 とすることもできる。

#### 【0023】

===ストレージ装置===

次に、本実施の形態に係るストレージ装置 600 のブロック図を図 2 に示す。なお図 2 には、ストレージ装置 600 が 5 台の情報処理装置 200 と通信可能に接続される場合の例を示す。

#### 【0024】

ストレージ装置 600 は、ストレージ制御装置 100 とストレージ駆動装置 300 とを備える。ストレージ制御装置 100 は、例えば情報処理装置 1 乃至 5 (200) から受信したコマンドに従ってストレージ駆動装置 300 に対する制御を行う。例えば情報処理装置 1 乃至 5 (200) からデータ入出力要求を受信して、ストレージ駆動装置 300 が備える記憶ボリューム 310 に対してデータの読み書きを行う。記憶ボリューム 310 とは、ハードディスクドライブ等の物理ディスクドライブにより提供される物理的な記憶領域である物理ボリュームと、物理ボリューム上に論理的に設定される論理的な記憶領域である論理ボリュームとを含む、データを記憶するための記憶領域をいう。

#### 【0025】

図 2 において、情報処理装置 1 乃至 4 (200) は SAN 500 を介してストレージ制御装置 100 と通信可能に接続されている。情報処理装置 5 (200) は、SAN 500 等のネットワークを介さずにストレージ制御装置 100 と接続されている。

#### 【0026】

===ストレージ駆動装置===

ストレージ駆動装置 300 はデータを記憶する複数のハードディスクドライブを備えている。これによりストレージ装置 600 は情報処理装置 1 乃至 5 (200) に対して大容量の記憶領域を提供することができる。これらのハードディスクドライブは、RAID (Redundant Arrays of Inexpensive Disks) を構成するようにすることもできる。

#### 【0027】

ストレージ制御装置 100 とストレージ駆動装置 300 との間は図 1 のようにネットワークを介さずに直接に接続される形態とすることもできるし、ネットワークを介して接続されるようにすることもできる。さらにストレージ駆動装置 300 はストレージ制御装置 100 と一体として構成されるようにすることもできる。

#### 【0028】

===ストレージ制御装置===

ストレージ制御装置 100 はチャンネル制御部 110、共有メモリ 120、キャッシュメモリ 130、ディスク制御部 140、管理端末 160、内部接続部 150 を備える。

#### 【0029】

チャンネル制御部 110 は、情報処理装置 200 や他のストレージ装置 600 との間で通信を行うための通信インタフェースを備える。またチャンネル制御部 110 は管理端末 160 と共に内部 LAN 151 で接続されている。これによりチャンネル制御部 110 に実行させるマイクロプログラム等を管理端末 160 から送信しインストールすることが可能となっている。チャンネル制御部 110 の構成については後述する。

#### 【0030】

内部接続部 150 はチャンネル制御部 110、共有メモリ 120、キャッシュメモリ 130、ディスク制御部 140 を相互に接続する。チャンネル制御部 110、共有メモリ 120、キャッシュメモリ 130、ディスク制御部 140 の間でのデータやコマンドの授受は内部接続部 150 を介することにより行われる。内部接続部 150 は例えばクロスバスイッチで構成される。

#### 【0031】

共有メモリ 120 及びキャッシュメモリ 130 は、チャンネル制御部 110 とディスク制御部 140 との間で授受されるコマンドやデータ等を記憶するメモリである。共有メモリ 120 は主に制御情報やコマンド等を記憶するために利用されるのに対し、キャッシュメモリ 130 は、主にデータを記憶するために利用される。

#### 【0032】

例えば、あるチャネル制御部 110 が情報処理装置 200 から受信したデータ入出力要求がデータ書き込み要求であった場合には、当該チャネル制御部 110 はデータ書き込み要求を共有メモリ 120 に書き込むと共に、情報処理装置 200 から受信した書き込みデータをキャッシュメモリ 130 に書き込む。一方、ディスク制御部 140 は共有メモリ 120 を監視しており、共有メモリ 120 にデータ書き込み要求が書き込まれたことを検出すると、当該データ書き込み要求に従ってキャッシュメモリ 130 から書き込みデータを読み出してストレージ駆動装置 300 に書き込む。

#### 【0033】

またあるチャネル制御部 110 が情報処理装置 200 から受信したデータ入出力要求がデータ読み出し要求であった場合には、読み出し対象となる読み出しデータがキャッシュメモリ 130 に存在するかどうかを調べる。ここでキャッシュメモリ 130 に存在すれば、チャネル制御部 110 はその読み出しデータを情報処理装置 200 に送信する。一方、読みだしデータがキャッシュメモリ 130 に存在しない場合には、当該チャネル制御部 110 はデータ読み出し要求を共有メモリ 120 に書き込むと共に、共有メモリ 120 を監視する。データ読み出し要求が共有メモリ 120 に書き込まれたことを検出したディスク制御部 140 は、ストレージ駆動装置 300 から読みだし対象となる読み出しデータを読み出してこれをキャッシュメモリ 130 に書き込むと共に、その旨を共有メモリ 120 に書き込む。そして、チャネル制御部 110 は読みだし対象となる読み出しデータがキャッシュメモリ 130 に書き込まれたことを検出すると、その読み出しデータを情報処理装置 200 に送信する。

#### 【0034】

このようにチャネル制御部 110 及びディスク制御部 140 の間では、キャッシュメモリ 130 を介してデータの授受が行われる。

#### 【0035】

なお、チャネル制御部 110 からディスク制御部 140 に対するデータの書き込みや読み出しの指示を共有メモリ 120 を介在させて間接的に行う構成の他、例えばチャネル制御部 110 からディスク制御部 140 に対してデータの書き込みや読み出しの指示を共有メモリ 120 を介さずに直接に行う構成とすることもできる。

また、チャネル制御部 110 とディスク制御部 140 とを一体的に構成し、両機能を合わせ持った制御部を設けるようにすることもできる。

#### 【0036】

本実施の形態に係る共有メモリ 120 には、図 9 に示すように、パリティグループ管理テーブル 710、パリティブロック管理テーブル 720、パリティブロック数総括管理テーブル 730、初期データ管理テーブル 740 が記憶される。これらの各テーブルについては後述する。

#### 【0037】

ディスク制御部 140 は、ストレージ駆動装置 300 と通信可能に接続され、データを記憶する記憶ボリューム 310 に記憶されるデータの読み書きを行う。例えば上述のように、チャネル制御部 110 が情報処理装置 200 から受信したデータ入出力要求に応じて、記憶ボリューム 310 に記憶されるデータの読み書きを行う。

#### 【0038】

各ディスク制御部 140 は管理端末 160 と共に内部 LAN 151 で接続されており、相互に通信を行うことが可能である。これにより、ディスク制御部 140 に実行させるマイクロプログラム等を管理端末 160 から送信しインストールすることが可能となっている。ディスク制御部 140 の構成については後述する。

#### 【0039】

本実施例においては、共有メモリ 120 及びキャッシュメモリ 130 がチャネル制御部 110 及びディスク制御部 140 に対して独立に設けられている場合について記載したが、本実施例はこの場合に限られるものでない。例えば共有メモリ 120 又はキャッシュメモリ 130 がチャネル制御部 110 及びディスク制御部 140 の各々に分散されて設けら

れることも好ましい。この場合、内部接続部150は、分散された共有メモリ120又はキャッシュメモリ130を有するチャンネル制御部110及びディスク制御部140を相互に接続させることになる。

#### 【0040】

また、チャンネル制御部110、ディスク制御部140、内部接続部150、共有メモリ120、キャッシュメモリ130の少なくともいずれかが一体として構成されているようにすることもできる。

#### 【0041】

===管理端末===

管理端末160はストレージ装置600を保守・管理するための情報機器である。オペレータは、管理端末160を操作することにより、例えばストレージ駆動装置300内のハードディスクドライブの構成の設定や、情報処理装置200とチャンネル制御部110との間の通信路であるパスの設定、記憶ボリューム310の設定、チャンネル制御部110やディスク制御部140において実行されるマイクロプログラムのインストール等を行うことができる。これらの設定や制御は、管理端末160が備えるユーザインタフェース、あるいは管理端末160で動作するWebサーバにより提供されるWebページを表示する情報処理装置1乃至5(200)等のユーザインタフェースから行うようにすることができる。

#### 【0042】

管理端末160はストレージ制御装置100に内蔵されている形態とすることもできるし、外付けされている形態とすることもできる。また管理端末160は、ストレージ装置600の保守・管理を専用に行うコンピュータとすることもできるし、汎用のコンピュータに保守・管理機能を持たせたものとすることもできる。

#### 【0043】

管理端末160の構成を示すブロック図を図5に示す。

管理端末160は、CPU161、メモリ162、ポート163、記録媒体読取装置164、入力装置165、出力装置166、記憶装置168を備える。

#### 【0044】

CPU161は管理端末160の全体の制御を司るもので、メモリ162に記憶された各種の動作を行うためのコードから構成されるストレージ管理プログラム162Aを実行することにより、ストレージ装置600の保守・管理機能を提供することができる。また同様に例えばストレージ管理プログラム162Aを実行することにより上記Webサーバとしての機能等を実現するようにすることができる。

#### 【0045】

記録媒体読取装置164は、記録媒体167に記録されているプログラムやデータを読み取るための装置である。読み取られたプログラムやデータはメモリ162や記憶装置168に格納される。従って、例えば記録媒体167に記録されたストレージ管理プログラム162Aを、記録媒体読取装置164を用いて上記記録媒体167から読み取って、メモリ162や記憶装置168に格納するようにすることができる。記録媒体167としてはフレキシブルディスクやCD-ROM、半導体メモリ等を用いることができる。記録媒体読取装置164は管理端末160に内蔵されている形態とすることもできるし、外付けされている形態とすることもできる。記憶装置168は、例えばハードディスク装置や半導体記憶装置等である。入力装置165はオペレータ等による管理端末160へのデータ入力等のために用いられるユーザインタフェースである。入力装置165としては例えばキーボードやマウス等が用いられる。出力装置166は情報を外部に出力するために用いられるユーザインタフェースである。出力装置166としては例えばディスプレイやプリンタ等が用いられる。ポート163は内部LAN151に接続されており、これにより管理端末160はチャンネル制御部110やディスク制御部140等と通信を行うことができる。またポート163は例えばLAN400等と通信可能に接続されるようにすることもできる。この場合管理端末160は、LAN400を通じて情報処理装置1乃至5(200

)と通信を行うようにすることもできる。

#### 【0046】

===外観図===

次に、本実施の形態に係るストレージ装置600の外観構成を図3に示す。また、ストレージ制御装置100の外観構成を図4に示す。

図3に示すように、本実施の形態に係るストレージ装置600はストレージ制御装置100及びストレージ駆動装置300がそれぞれの筐体に納められた形態をしている。図3に示す例では、ストレージ制御装置100の筐体の両側にストレージ駆動装置300の筐体が配置されている。

#### 【0047】

ストレージ制御装置100は、正面中央部に管理端末160が備えられている。管理端末160はカバーで覆われており、図4に示すようにカバーを開けることにより管理端末160を使用することができる。なお図4に示した管理端末160はいわゆるノート型パーソナルコンピュータの形態をしているが、どのような形態とすることも可能である。

#### 【0048】

管理端末160の下部には、チャンネル制御部110やディスク制御部140、キャッシュメモリ130、共有メモリ120、内部接続部150を装着するためのスロットが設けられている。チャンネル制御部110やディスク制御部140、キャッシュメモリ130、共有メモリ120、内部接続部150は回路基板を備えてボードとして構成されており、これらのボードが各スロットに装着される。各スロットにはこれらのボードを装着するためのガイドレールが設けられている。ガイドレールに沿って各ボードをスロットに挿入することにより、チャンネル制御部110やディスク制御部140、キャッシュメモリ130、共有メモリ120、内部接続部150をストレージ制御装置100に装着することができる。各スロットの奥手方向正面部には、各ボードをストレージ制御装置100と電氣的に接続するためのコネクタが設けられている。

#### 【0049】

またストレージ制御装置100には、チャンネル制御部110等から発生する熱を放出するためのファン170が設けられている。ファン170はストレージ制御装置100の上面部に設けられる他、スロットの上部にも設けられている。

#### 【0050】

===チャンネル制御部===

チャンネル制御部110の構成を図6に示す。

チャンネル制御部110は回路基板を備えた一つのユニット化されたボードとして構成される。チャンネル制御部110は一枚もしくは複数枚の回路基板を含んで構成される。回路基板には、インタフェース部111、メモリ113、CPU112、NVRAM (nonvolatile random-access memory) 114、コネクタ115が形成される。

#### 【0051】

インタフェース部111は、情報処理装置200や他のストレージ装置600との間で通信を行うための通信インタフェースや、内部接続部150を介して共有メモリ120やキャッシュメモリ130、チャンネル制御部110等との間で通信を行うための通信インタフェースを備える。

CPU112は、チャンネル制御部110全体の制御を司る。CPU112によりメモリ113やNVRAM114に格納された各種プログラムが実行されることにより本実施の形態に係るチャンネル制御部110の機能が実現される。

NVRAM114はCPU112の制御を司るプログラムを格納する不揮発性メモリである。NVRAM114に記憶されるプログラムの内容は、管理端末160からの指示により書き込みや書き換えを行うことができる。

チャンネル制御部110はコネクタ115を備えている。コネクタ115がストレージ制御装置100側のコネクタと嵌合することにより、チャンネル制御部110はストレージ制御装置100の内部接続部150や管理端末160等と電氣的に接続される。



メモリ 113 には、ストレージ制御プログラム 810、データ送受信制御プログラム 820、パリティ制御プログラム 830、障害検知プログラム 840 が記憶される。

【0052】

ストレージ制御プログラム 810 は、記憶ボリューム 310 に記憶されるデータの出入力制御やストレージ装置 600 における各種設定のための制御等を行うためのプログラムである。例えば後述するパリティグループ管理テーブル 710、パリティブロック管理テーブル 720、パリティブロック数統括管理テーブル 730、初期データ管理テーブル 740 の作成更新等は、CPU 112 によりストレージ制御プログラム 810 が実行されることにより行われる。また第 1 乃至第 4 の記憶制御部、及び第 1 乃至第 5 のデータ記憶制御部は、例えばチャネル制御部 110 において CPU 112 によりストレージ制御プログラム 810 が実行され、ディスク制御部 140 や共有メモリ 120、キャッシュメモリ 130、ディスク駆動装置 300 等と協働して動作することにより実現される。

【0053】

データ送受信制御プログラム 820 は、他のストレージ装置 600 や情報処理装置 200 との間のデータの送受信を制御するためのプログラムである。従って第 1 乃至第 6 の受信部や第 1 乃至第 2 の送信部、第 1 乃至第 7 のデータ受信部、第 1 乃至第 8 のデータ送信部は、例えば CPU 112 によりデータ送受信制御プログラム 820 が実行されることにより実現される。

【0054】

パリティ制御プログラム 830 は、排他的論理和の演算を行うためのプログラムである。例えば複数の他のストレージ装置 600 からそれぞれ受信したデータの排他的論理和の演算や、記憶ボリューム 310 に記憶されているデータと情報処理装置 200 から受信したデータとの排他的論理和の演算、記憶ボリューム 310 に記憶されているデータと他のストレージ装置 600 から受信したデータとの排他的論理和の演算などを行う。従って第 1 乃至第 5 の演算制御部や第 1 乃至第 7 のデータ演算制御部は、例えば CPU 112 によりパリティ制御プログラム 830 が実行されることにより実現される。

【0055】

障害検知プログラム 840 は、ストレージ装置 600 に発生する障害を検知するためのプログラムである。例えば記憶ボリューム 310 に記憶されているデータに異常がないかを検知する。

【0056】

記憶ボリューム 310 に記憶されているデータに異常を検知したストレージ装置 600 が、図 1 におけるストレージ装置 1 乃至 3 (600) のいずれかの場合、例えばストレージ装置 1 (600) の場合には、ストレージ装置 1 (600) はストレージ装置 4 (600) に対して元のデータの送信要求を送信する。そうするとストレージ装置 4 (600) は、ストレージ装置 1 (600) 以外のストレージ装置 600 すなわちストレージ装置 2 乃至 3 (600) に対して、それぞれのストレージ装置 600 に記憶されているデータの送信要求を送信する。そしてストレージ装置 4 (600) は、ストレージ装置 2 乃至 4 (600) に記憶されているデータの排他的論理和を演算することにより、ストレージ装置 1 (600) に記憶されていた元のデータを復元し、ストレージ装置 1 (600) に送信する。

【0057】

一方、記憶ボリューム 310 に記憶されているデータに異常を検知したストレージ装置 600 が、図 1 におけるストレージ装置 4 (600) の場合には、ストレージ装置 1 乃至 3 (600) に対して、それぞれのストレージ装置 600 に記憶されているデータの送信要求を送信する。そしてストレージ装置 4 (600) は、ストレージ装置 1 乃至 3 (600) に記憶されているデータの排他的論理和を演算することにより、ストレージ装置 4 (600) に記憶されていた元のデータを復元する。

【0058】

なお、ストレージ制御プログラム 810 や、データ送受信制御プログラム 820、パリ

ティ制御プログラム 830、障害検知プログラム 840 は、それぞれ個別のプログラムとすることもできるし、これらのプログラムの少なくとも一部が同一プログラムにより構成されるようにすることもできる。また各プログラムが複数のプログラムにより構成されるようにすることもできる。

#### 【0059】

=== ディスク制御部 ===

次にディスク制御部 140 の構成を示す図を図 7 に示す。

ディスク制御部 140 は、回路基板を備えた一つのユニット化されたボードとして構成される。ディスク制御部 140 は一枚もしくは複数枚の回路基板を含んで構成される。回路基板には、インタフェース部 141、メモリ 143、CPU 142、NVRAM 144、コネクタ 145 が形成される。

#### 【0060】

インタフェース部 141 は、内部接続部 150 を介してチャネル制御部 110 等との間で通信を行うための通信インタフェースや、ストレージ駆動装置 300 との間で通信を行うための通信インタフェースを備えている。

CPU 142 は、ディスク制御部 140 全体の制御を司る。CPU 142 によりメモリ 143 や NVRAM 144 に格納された各種プログラムが実行されることにより本実施の形態に係るディスク制御部 140 の機能が実現される。

NVRAM 144 は CPU 142 の制御を司るプログラムを格納する不揮発性メモリである。NVRAM 144 に記憶されるプログラムの内容は、管理端末 160 からの指示により書き込みや書き換えを行うことができる。

またディスク制御部 140 はコネクタ 145 を備えている。コネクタ 145 がストレージ制御装置 100 側のコネクタと嵌合することにより、ディスク制御部 140 はストレージ制御装置 100 の内部接続部 150 や、ストレージ駆動装置 300、管理端末 160 等と電氣的に接続される。

#### 【0061】

なお、上述のチャネル制御部 110 において実行されるパリティ制御プログラム 830 や障害検知プログラム 840 がディスク制御部 140 のメモリ 143 に記憶されるようにすることもできる。この場合、CPU 142 によりパリティ制御プログラム 830 や障害検知プログラム 840 が実行されることにより、ディスク制御部 140 において上記排他的論理和の演算やストレージ装置 600 に発生する障害の検知が行われるようにすることもできる。

#### 【0062】

=== 情報処理装置 ===

次に本実施の形態に係る情報処理装置 200 の構成を示すブロック図を図 8 に示す。

情報処理装置 200 は、CPU 210、メモリ 220、ポート 230、記録媒体読取装置 240、入力装置 250、出力装置 260、記憶装置 280 を備える。

#### 【0063】

CPU 210 は情報処理装置 200 の全体の制御を司るもので、メモリ 220 に記憶された各種の動作を行うためのコードから構成されるプログラム 220A を実行することにより本実施の形態に係る各種機能を実現する。例えば、上述した銀行の自動預金預け払いサービス等の情報処理サービスの提供は、CPU 210 がプログラム 220A を実行することにより行われる。また、CPU 210 がプログラム 220A を実行することにより、上述した管理端末 160 で動作する Web サーバにより提供される Web ページの表示や、ハードディスクドライブの構成の変更や、情報処理装置 200 とチャネル制御部 110 との間の通信路であるパスの設定、記憶ボリューム 310 の設定等を行うことができる。記録媒体読取装置 240 は記録媒体 270 に記録されているプログラムやデータを読み取るための装置である。読み取られたプログラムやデータはメモリ 220 や記憶装置 280 に格納される。従って、例えば記録媒体 270 に記録されたプログラム 220A を、記録媒体読取装置 240 を用いて上記記録媒体 270 から読み取って、メモリ 220 や記憶装

置 280 に記憶するようにすることができる。

記録媒体 270 としてはフレキシブルディスクや CD-ROM、半導体メモリ等を用いることができる。記録媒体読取装置 240 は情報処理装置 200 に内蔵されている形態とすることもできるし、外付されている形態とすることもできる。

#### 【0064】

記憶装置 280 は、例えばハードディスク装置や半導体記憶装置等とすることができる。また記憶装置 280 は情報処理装置 200 に内蔵されるようにすることもできるし、外付けされるようにすることもできる。外付けされる場合には、通信ネットワークを介して接続される他の情報処理装置 200 の記憶装置 280 とすることもできる。またストレージ装置 600 とすることもできる。入力装置 250 は情報処理装置 200 を操作するオペレータ等による情報処理装置 200 へのデータ入力等のために用いられるユーザインタフェースである。入力装置 250 としては例えばキーボードやマウス等が用いられる。出力装置 260 は情報を外部に出力するためのユーザインタフェースである。出力装置 260 としては例えばディスプレイやプリンタ等が用いられる。ポート 230 は、チャンネル制御部 110 と通信を行うための装置である。またポート 230 は、LAN400 等の通信ネットワークを通じて他の情報処理装置 200 や管理端末 160 と通信を行うことができるようにすることもできる。この場合、例えばプログラム 220A をポート 230 を介して他の情報処理装置 200 から受信して、メモリ 220 や記憶装置 280 に記憶するようにすることもできる。

#### 【0065】

===バックアップデータの作成===

上述したように、本実施の形態においてはストレージ装置 1 乃至 3（データストレージ装置、他のストレージ装置、第 1 のストレージ装置）（600）に記憶されるデータをバックアップするためのデータがストレージ装置 4（パリティストレージ装置、第 2 のストレージ装置）（600）に記憶される。そこでまず、ストレージ装置 4（600）にストレージ装置 1 乃至 3（600）に記憶されるデータをバックアップするためのデータを記憶する際の処理の流れについて、図 14 乃至図 26 を用いて説明する。

#### 【0066】

データストレージ装置 600 に記憶されるデータをバックアップするためのデータをパリティストレージ装置 600 に記憶する場合のやり方として、同期方式、又は非同期方式を採用することができる。

#### 【0067】

同期方式とは、パリティストレージ装置 600 に通信可能に接続される全てのデータストレージ装置 600 に記憶されるデータを用いて排他的論理和を演算し、その演算結果であるパリティデータをパリティストレージ装置 600 に記憶する方式をいう。従って、例えばパリティストレージ装置 600 に通信可能に接続されるデータストレージ装置 600 が追加される場合には、同期方式の場合には、追加されるストレージ装置 600 を含む全てのデータストレージ装置 600 に記憶されるデータを用いて再度排他的論理和が演算され、パリティストレージ装置 600 に記憶される。

#### 【0068】

これを式で表せば以下のようなになる。すなわち例えば、データストレージ装置 600 がストレージ装置 1（600）とストレージ装置 2（600）との 2 台の場合に、データストレージ装置としてストレージ装置 3（600）が追加される場合には、ストレージ装置 1（600）に記憶されるデータを D1、ストレージ装置 2（600）に記憶されるデータを D2、ストレージ装置 3（600）に記憶されるデータを D3 とすれば、パリティストレージ装置 600 に記憶されるデータは、ストレージ装置 3（600）が追加される際に、 $D1 \oplus D2 \oplus D3$  が演算されることにより作成される。

#### 【0069】

一方、非同期方式とは、パリティストレージ装置 600 に通信可能に接続されるデータストレージ装置 600 が追加される場合に、パリティストレージ装置 600 に元々記憶さ

れるデータと追加されるストレージ装置 600 に記憶されるデータとを用いて排他的論理和が演算され、パリティストレージ装置 600 に記憶される方式をいう。

#### 【0070】

同様にこれを式で表せば以下のようなになる。すなわちストレージ装置 3 (600) が追加される前にパリティストレージ装置 600 に記憶されていたデータを D4 とすれば、ストレージ装置 3 (600) が追加される際に、D4 (EXOR) D3 が演算されることによりパリティストレージ装置 600 に記憶されるデータが作成される。

#### 【0071】

まず図 14 に、同期方式の場合の処理の流れを、パリティストレージ装置 600 における処理と、データストレージ装置 600 における処理とを分けて示す。

まずパリティストレージ装置 600 は設定入力情報を受け付ける (S1000)。設定入力情報の受付は、例えば図 15 に示すようなウインドウ画面を、管理端末 160 が備える出力装置 166 に表示し、オペレータ等からの入力を入力装置 166 から受け付けることにより行うことができる。オペレータ等から受け付ける設定内容は、例えば「パリティブロック長」、「パリティグループ構成ストレージ装置」、「パリティデータ記憶ストレージ装置」、「初期化方式」とすることができる。

#### 【0072】

「パリティブロック長」欄は、パリティブロックの大きさを定義するための欄である。パリティブロックとは、ストレージ装置 600 が備えるハードディスクドライブのデータ記憶領域を論理的に区分してなる記憶領域 (記憶ブロック) を言う。パリティブロック長はパリティブロックのサイズである。図 15 には、パリティブロック長を 512 バイトとする場合の例が示される。各パリティブロックは、ストレージ装置 600 内において一意に付与される識別子 (パリティブロック番号) により特定される。

#### 【0073】

「パリティグループ構成ストレージ装置」欄は、データストレージ装置 600 を構成するストレージ装置 600 を定義するための欄である。図 15 には、データストレージ装置 600 として、ストレージ装置 1 乃至 3 (600) を定義する場合の例が示される。

#### 【0074】

「パリティデータ記憶ストレージ装置」欄は、データストレージ装置 600 に記憶されるデータをバックアップするためのデータを記憶するパリティストレージ装置 600 を構成するストレージ装置 600 を定義するための欄である。図 15 には、パリティストレージ装置 600 として、ストレージ装置 4 (600) を定義する場合の例が示される。

#### 【0075】

「初期化方式」欄は、データストレージ装置 600 からパリティストレージ装置 600 へのデータの送信のやり方を定義するための欄である。「初期化方式」には「順次式」と「ランダム式」とがある。「順次式」の場合は、データストレージ装置 600 の各パリティブロックに記憶されるデータがパリティブロック番号順に、パリティストレージ装置 600 に送信される。「ランダム式」の場合には、データストレージ装置 600 が情報処理装置 200 からデータ書き込み要求を受信するのを契機に、データ書き込み要求の対象となったパリティブロックに記憶されるデータと書き込みデータとの排他的論理和の演算結果がパリティストレージ装置 600 に送信される。なお「ランダム式」において、データストレージ装置 600 が情報処理装置 200 からデータ読み出し要求を受信するのを契機に、データ読み出し要求の対象となったパリティブロックに記憶されるデータがパリティストレージ装置 600 に送信されるようにすることもできる。

#### 【0076】

図 15 の画面において、マウス等により「OK」欄をクリックすることにより、これらの設定入力情報がストレージ装置 600 に入力される。

そうするとパリティストレージ装置 600 は、パリティブロック長を用いてパリティブロック数を求める (S1001)。具体的には、パリティストレージ装置 600 が備える記憶領域の記憶容量をパリティブロック長で割って求めた値がパリティブロック数となる。

**【0077】**

そしてパリティストレージ装置600は、パリティグループ管理テーブル710とパリティブロック管理テーブル720とを作成する(S1002)。パリティグループ管理テーブル710を図10に示す。またパリティブロック管理テーブル720を図11に示す。

**【0078】**

パリティグループ管理テーブル710は、「データ記憶ストレージ装置」欄と、「パリティ記憶ストレージ装置」欄とを備える。「データ記憶ストレージ装置」欄にはデータストレージ装置600が記憶される。「パリティ記憶ストレージ装置」欄にはパリティストレージ装置600が記憶される。上述したように、パリティグループ管理テーブル710は共有メモリ120に記憶される。

**【0079】**

一方パリティブロック管理テーブル720は、「パリティブロック長」欄と、「パリティブロック数」欄と、「論理ボリューム番号」欄とを備える。「パリティブロック長」欄にはパリティブロック長が記憶される。「パリティブロック数」欄にはパリティブロック数が記憶される。「論理ボリューム番号」欄には、ストレージ装置600が備える各論理ボリュームの各記憶領域の先頭のパリティブロック番号が記憶される。上述したように、パリティブロック管理テーブル720も共有メモリ120に記憶される。

**【0080】**

続いてパリティストレージ装置600は、各データストレージ装置600へこれらの設定入力情報を送信する(S1003)。

そうすると各データストレージ装置600は、パリティストレージ装置600から送信されたパリティブロック長を用いてパリティブロック数を求め(S1004)、パリティストレージ装置600と同様に、パリティグループ管理テーブル710とパリティブロック管理テーブル720とを作成する(S1005)。

**【0081】**

このように、パリティストレージ装置600とデータストレージ装置600とで、それぞれの記憶領域をパリティデータ長毎に論理的に区分することにより、両ストレージ装置600の記憶領域をパリティブロック毎に対応付けることが可能となる。例えばパリティストレージ装置600とデータストレージ装置600とで同一のパリティブロック番号で特定されるパリティブロック同士を対応づけることが可能となる。その様子を示したのが図16である。図16においては、パリティブロック番号を特定(パリティブロック番号12356)することにより、パリティストレージ装置600とデータストレージ装置600とのそれぞれのパリティブロックを特定することができることが示される。

このようにすることにより、記憶ボリューム310の制御の方式が異なるストレージ装置600を用いてストレージシステムを構成する場合にも、データストレージ装置600に記憶されるデータをバックアップすることができるようになる。

**【0082】**

なお記憶ボリューム310の制御の方式としては、例えば、メインフレーム系情報処理システムで使用されることの多いCKD(Count Key Data)方式や、オープン系情報処理システムで使用されることの多いFBA(Fixed Block Architecture)方式等がある。

**【0083】**

次に各データストレージ装置600は、パリティストレージ装置600へ、S1004にて求めたパリティブロック数を送信する(S1006)。

**【0084】**

パリティストレージ装置600は、各データストレージ装置600からそれぞれのパリティブロック数を受信すると、S1001で求めたパリティブロック数と比較し、パリティ生成可否を判定する(S1007)。S1001で求めたパリティブロック数よりも大きなパリティブロック数をデータストレージ装置600から受信した場合には、S1007において“否”に進み、同期初期コピー不可の旨を管理端末160の出力装置166に表示して処理を終了する(S1015)。なぜならばこの場合、データストレージ装置600の各パリティブロッ

クに記憶されるデータのうち、パリティストレージ装置 600 のパリティブロック番号（パリティブロックの識別子）の最大値よりも大きなパリティブロック番号で特定されるパリティブロックに記憶されるデータをバックアップするためのデータをパリティストレージ装置 600 に記憶することはできないからである。

#### 【0085】

パリティストレージ装置 600 は、S1001 で求めたパリティブロック数よりも大きなパリティブロック数をデータストレージ装置 600 から受信しない場合には、S1007 において“可”に進み、パリティブロック数統括管理テーブル 730 を作成する（S1008）。パリティブロック数統括管理テーブル 730 を図 12 に示す。図 12 に示すように、パリティブロック数統括管理テーブル 730 には各ストレージ装置 600 のパリティブロック数が記憶される。なお、パリティストレージ装置 600 にて S1007 におけるパリティ生成可否が判定される様子を示す図を図 17 に示す。

#### 【0086】

続いてパリティストレージ装置 600 は、各データストレージ装置 600 へ、各パリティブロックに記憶されるデータをパリティブロック番号順に先頭から順次送信するように要求を送る（S1009）。そうすると各データストレージ装置 600 は、要求に従って、各パリティブロックに記憶されるデータ（第 1 の記憶データ）の複製を先頭から順次、パリティブロック番号（第 1 の識別子）と共に送信する（S1010）。

#### 【0087】

ここで各ストレージ装置 600 間でデータを授受する際のデータフォーマットを図 33 に示す。データフォーマットは、「パリティブロック番号」欄と、「ステータス情報」欄と、「データ」欄とを含む。「パリティブロック番号」欄には、授受されるデータのパリティブロック番号が記載される。「ステータス情報」欄には各種制御情報が記載される。例えば、データの送信先ストレージ装置 600 を示す情報や、送信元ストレージ装置 600 を示す情報、あるいは、このデータフォーマットにより送信されるデータがパリティデータを作成するために用いられるデータである旨の情報、後述する情報処理装置 200 により更新されたデータである旨の情報、「データ」欄に記載されるデータのデータ長すなわちパリティブロック長等が記載される。「データ」欄は、授受されるデータが記載される。

#### 【0088】

パリティストレージ装置 600 は、データストレージ装置 600 のそれぞれから、パリティブロックのそれぞれに記憶されるデータの複製を、各パリティブロックを特定するパリティブロック番号（第 1 の識別子）と共に受信する。そしてパリティストレージ装置 600 は、データストレージ装置 600 のそれぞれから受信した各パリティブロックのデータの複製のうち、パリティブロック番号が同一のデータの複製の排他的論理和を演算する（S1011）。その後、パリティストレージ装置 600 のパリティブロックのそれぞれを特定するパリティブロック番号（第 2 の識別子）が、データストレージ装置 600 から受信したパリティブロック番号と同一のパリティブロックに、排他的論理和の演算結果を記憶する（S1012）。

#### 【0089】

そしてパリティストレージ装置 600 は初期データ管理テーブル 740 を更新する（S1013）。初期データ管理テーブル 740 は、各データストレージ装置 600 に記憶されるデータのバックアップがどこまで行われているかを示すためのテーブルである。初期データ管理テーブル 740 の例を図 13 に示す。図 13（A）は、データストレージ装置 600 単位で、各データストレージ装置 600 に記憶されるデータのバックアップがどこまで行われているかを示す場合の例である。

すなわち、「再計算中」欄が“OFF”のデータストレージ装置 600 は、すでにデータのバックアップのための計算、すなわち排他的論理和の演算が完了していることを示す。「再計算中」欄が“ON”のデータストレージ装置 600 は、未だデータのバックアップのための計算、すなわち排他的論理和の演算が完了していないことを示す。同期方式で

バックアップデータの作成が行われる場合には、図13 (A) に示す初期データ管理テーブル740の「再計算中」欄は、全てのデータストレージ装置600について”ON”であるか、全てのデータストレージ装置600について”OFF”であるかのいずれかになる。図13 (B) は、排他的論理和の演算のためのデータの送信が順方式行われる場合に、「再計算中」欄が”ON”のデータストレージ装置600に対して、どのパリティブロックまで排他的論理和の演算が完了しているかを示す「再計算位置」欄を備える場合の例である。図13 (C) は、排他的論理和の演算のためのデータの送信がランダム式で行われる場合に、「再計算中」欄が”ON”のデータストレージ装置600に対して、どのパリティブロックの排他的論理和の演算が完了しているかを示すための「再計算終了」欄を備える場合の例である。初期データ管理テーブル740としては、適宜、図13の (A)、(B)、又は (C) を採用することができる。

#### 【0090】

図13 (A) を採用した場合の初期データ管理テーブル740の更新の様子を図18に示す。図13 (B) を採用した場合の初期データ管理テーブル740の更新の様子を図19に示す。図13 (C) を採用した場合の初期データ管理テーブル740の更新の様子を図20に示す。

#### 【0091】

最後にパリティストレージ装置600は、初期データ管理テーブル740を参照し、全てのパリティブロックについての排他的論理和の演算結果を記憶したか否かを確認し (S1014) 、”Yes” の場合は処理を終了する。

#### 【0092】

以上の処理により、パリティストレージ装置600の各パリティブロックには、各データストレージ装置600のそれぞれ対応する各パリティブロックのデータにより演算された排他的論理和の演算結果を記憶することができる。なお、ここではパリティ生成可否の判定をパリティストレージ装置600で行ったが (S1007) 、各データストレージ装置600で行うようにすることもできる。その場合の処理の流れを図21に示す。

#### 【0093】

まずパリティストレージ装置600は設定入力情報を受け付ける (S2000) 。

そうするとパリティストレージ装置600は、パリティブロック長を用いてパリティブロック数を求める (S2001) 。

そしてパリティストレージ装置600は、パリティグループ管理テーブル710とパリティブロック管理テーブル720とを作成する (S2002) 。

続いてパリティストレージ装置600は、各データストレージ装置600へこれらの設定入力情報と、S2001にて求めたパリティブロック数とを送信する (S2003) 。

そうすると各データストレージ装置600は、パリティストレージ装置600から送信されたパリティブロック長を用いてパリティブロック数を求め (S2004) 、パリティグループ管理テーブル710とパリティブロック管理テーブル720とを作成する (S2005) 。

#### 【0094】

各データストレージ装置600は、パリティストレージ装置600から送信されたパリティブロック数と、S2004にて求めたパリティブロック数とを比較し (S2006) 、パリティ生成可否を判定する。パリティストレージ装置600から送信されたパリティブロック数の方がS2004にて求めたパリティブロック数よりも大きい場合には、パリティ生成可能と判定し、パリティストレージ装置600から送信されたパリティブロック数の方がS2004にて求めたパリティブロック数よりも小さい場合には、パリティ生成不可と判定する。そしてその判定の結果をパリティストレージ装置600へ送信する (S2007) 。

#### 【0095】

パリティストレージ装置600は、各データストレージ装置600からそれぞれのパリティ生成可否の判定結果を受信し、一つでもパリティ生成不可の判定結果があった場合には、S2008において”否”に進み、同期初期コピー不可の旨を管理端末160の出力装置

166に表示して処理を終了する(S2016)。一方、全てパリティ生成可能の判定であった場合には、“可”に進み、パリティブロック数統括管理テーブル730を作成する(S2009)。

#### 【0096】

続いてパリティストレージ装置600は、各データストレージ装置600へ、各パリティブロックに記憶されるデータをパリティブロック番号順に先頭から順次送信するように要求を送る(S2010)。そうすると各データストレージ装置600は、要求に従って、各パリティブロックに記憶されるデータの複製を先頭から順次、パリティブロック番号と共に送信する(S2011)。

#### 【0097】

そしてパリティストレージ装置600は、データストレージ装置600のそれぞれから、パリティブロックのそれぞれに記憶されるデータの複製を、各パリティブロックを特定するパリティブロック番号と共に受信し、データストレージ装置600のそれぞれから受信した各パリティブロックのデータの複製のうち、パリティブロック番号が同一のデータの複製の排他的論理和を演算する(S2012)。その後、パリティストレージ装置600のパリティブロックのそれぞれを特定するパリティブロック番号が、データストレージ装置600から受信したパリティブロック番号と同一のパリティブロックに、排他的論理和の演算結果を記憶する(S2013)。

そしてパリティストレージ装置600は初期データ管理テーブル740を更新する(S2014)。

最後にパリティストレージ装置600は、初期データ管理テーブル740を参照し、全てのパリティブロックについての排他的論理和の演算結果を記憶したか否かを確認し(S2015)、“Yes”の場合は処理を終了する。

各データストレージ装置600にてパリティ生成可否が判定される場合の処理の様子を図22に示す。

#### 【0098】

次に、図15の同期方式パリティグループ設定画面において、「初期化方式」欄に“ランダム式”が入力される場合の処理の流れを図23に示す。

まずパリティストレージ装置600は設定入力情報を受け付ける(S3000)。

パリティストレージ装置600は、入力されたパリティブロック長を用いてパリティブロック数を求める(S3001)。

そしてパリティストレージ装置600は、パリティグループ管理テーブル710とパリティブロック管理テーブル720とを作成する(S3002)。

#### 【0099】

続いてパリティストレージ装置600は、各データストレージ装置600へこれらの設定入力情報を送信する(S3003)。

そうすると各データストレージ装置600は、パリティストレージ装置600から送信されたパリティブロック長を用いてパリティブロック数を求め(S3004)、パリティグループ管理テーブル710とパリティブロック管理テーブル720とを作成する(S3005)。

各データストレージ装置600は、パリティストレージ装置600へ、S3004にて求めたパリティブロック数を送信する(S3006)。

パリティストレージ装置600は、各データストレージ装置600からそれぞれのパリティブロック数を受信すると、S3001で求めたパリティブロック数と比較し、パリティ生成可否を判定する(S3007)。S3001で求めたパリティブロック数よりも大きなパリティブロック数をデータストレージ装置600から受信した場合には、S3007において“否”に進み、同期初期コピー不可の旨を管理端末160の出力装置166に表示して処理を終了する(S3009)。

#### 【0100】

パリティストレージ装置600は、S3001で求めたパリティブロック数よりも大きなパ



リティブロック数をデータストレージ装置 600 から受信しない場合には、S3007において”可”に進み、パリティブロック数統括管理テーブル 730 を作成する (S3008)。

その後は、パリティストレージ装置 600 は、情報処理装置 200 からデータ書き込み要求を受信したデータストレージ装置 600 から、データ書き込み要求の対象となったパリティブロックに記憶されるデータと書き込みデータとの排他的論理和の演算結果を、そのデータ書き込み要求の対象となったパリティブロック番号と共に受信し、同一パリティブロック番号の演算結果が各データストレージ装置 600 から送信される毎にそれらの排他的論理和を演算し、対応するパリティブロックにその排他的論理和の演算結果を記憶する。

#### 【0101】

このようにすることにより、パリティストレージ装置 600 の各パリティブロックには、各データストレージ装置 600 のそれぞれ対応するパリティブロックのデータにより演算された排他的論理和の演算結果を記憶することができる。

#### 【0102】

次に、非同期方式によりデータストレージ装置 600 に記憶されるデータをバックアップするためのデータをパリティストレージ装置 600 に記憶する場合の処理の流れについて図 24 を用いて説明する。

#### 【0103】

まずパリティストレージ装置 600 は設定入力情報を受け付ける (S4000)。非同期方式の場合の設定入力情報の受付は、例えば図 25 に示すようなウインドウ画面を、管理端末 160 が備える出力装置 166 に表示し、オペレータ等からの入力を入力装置 166 から受け付けることにより行うことができる。オペレータ等から受け付ける設定内容は、例えば「追加するデータストレージ装置」、「初期化方式」、「削除するデータストレージ装置」とすることができる。

#### 【0104】

図 25 に示すウインドウ画面においては、「現在の構成」欄に、現在のストレージシステムの構成が示される。すなわち、パリティストレージ装置 600 はストレージ装置 4 (600) であり、データストレージ装置 600 はストレージ装置 1 (600) とストレージ装置 2 (600) であり、パリティブロック長が 512 バイトであることが示される。この場合、ストレージ装置 4 (600) のハードディスクドライブの各パリティブロックには、ストレージ装置 1 (600) とストレージ装置 2 (600) とのそれぞれ対応するパリティブロックに記憶されるデータから演算された排他的論理和の全ての演算結果が記憶されている。図 25 には、このような「現在の構成」に対して、ストレージ装置 3 (600) をデータストレージ装置 600 として追加する場合の例が示される。もちろん、図 25 に示すウインドウ画面は一例であり、例えばパリティストレージ装置 600 を変更するような入力を行えるようにすることもできるし、パリティブロック長を変更するような入力を行えるようにすることも可能である。

#### 【0105】

図 25 の画面において、マウス等により”OK”欄をクリックすることにより、これらの設定入力情報がストレージ装置 600 に入力される。なお、図 25 には「初期化方式」欄が”ランダム方式”となっているが、図 24 に示すフローチャートは”順次式”の場合の例を示す。”ランダム式”の場合のフローチャートは図 26 に示す。また「削除するデータストレージ装置」欄にデータストレージ装置 600 が記載された場合のフローチャートは、図 32 に示す。

#### 【0106】

そうするとパリティストレージ装置 600 は、まず「追加するデータストレージ装置」欄に記載されたデータストレージ装置 600 が最初のストレージ装置 600 であるか否かを判定する (S4001)。最初のストレージ装置 600 とは、最初のデータストレージ装置 600 をいう。つまり図 25 の非同期式パリティグループ設定画面の「現在の構成」欄の「データストレージ装置」欄にデータストレージ装置 600 が記載されていない場合には

、「追加するデータストレージ装置」欄に記載されたストレージ装置 600 は最初のストレージ装置 600 となる。最初のストレージ装置 600 でない場合には S4001 において”No”に進む。

【0107】

また最初のストレージ装置 600 である場合には S4001 において”Yes”に進む。そしてパリティストレージ装置 600 は、パリティブロック長を用いてパリティブロック数を求める (S4002)。

そしてパリティストレージ装置 600 は、パリティグループ管理テーブル 710 とパリティブロック管理テーブル 720 とを作成する (S4003)。

続いてパリティストレージ装置 600 は、追加されるデータストレージ装置 600 へパリティブロック長を送信する (S4004)。

そうすると追加されるデータストレージ装置 600 は、パリティストレージ装置 600 から送信されたパリティブロック長を用いてパリティブロック数を求め (S4005)、パリティグループ管理テーブル 710 とパリティブロック管理テーブル 720 とを作成する (S4006)。

追加されるデータストレージ装置 600 は、パリティストレージ装置 600 へ、S4005 にて求めたパリティブロック数を送信する (S4007)。

【0108】

パリティストレージ装置 600 は、追加されるデータストレージ装置 600 からパリティブロック数を受信すると、S4002 で求めたパリティブロック数と比較し、パリティ生成可否を判定する (S4008)。S4002 で求めたパリティブロック数よりも大きなパリティブロック数を追加されるデータストレージ装置 600 から受信した場合には、S4008 において”否”に進み、データストレージ装置 600 の追加が不可である旨を管理端末 160 の出力装置 166 に表示して処理を終了する (S4016)。

【0109】

パリティストレージ装置 600 は、S4002 で求めたパリティブロック数よりも大きなパリティブロック数を追加されるデータストレージ装置 600 から受信しない場合には、S4008 において”可”に進み、パリティブロック数統括管理テーブル 730 を作成する (S4009)。

【0110】

続いてパリティストレージ装置 600 は、追加されるデータストレージ装置 600 へ、各パリティブロックに記憶されるデータ (第 1 の記憶データ) の複製をパリティブロック番号と共にパリティブロック番号順に先頭から順次送信させるための送信要求を送信する (S4010)。そうすると追加されるデータストレージ装置 600 は、送信要求に応じて、各パリティブロックに記憶されるデータの複製を先頭から順次、パリティブロック番号と共にパリティストレージ装置 600 に送信する (S4011)。

【0111】

そしてパリティストレージ装置 600 は、追加されるデータストレージ装置 600 から、パリティブロックのそれぞれに記憶されるデータの複製を、パリティブロック番号と共に受信し、追加されたデータストレージ装置 600 から受信した各パリティブロックのデータの複製と、そのパリティブロック番号と同一のパリティブロック番号で特定されるパリティストレージ装置 600 のパリティブロックのデータ (旧パリティデータ、第 2 の記憶データ) との排他的論理和を演算し (S4012)、その演算結果 (新パリティデータ) を旧パリティデータが記憶されていたパリティストレージ装置 600 のパリティブロックに記憶する (S4013)。

【0112】

そしてパリティストレージ装置 600 は初期データ管理テーブル 740 を更新する (S4014)。

最後にパリティストレージ装置 600 は、初期データ管理テーブル 740 を参照して、全てのパリティブロックについての排他的論理和の演算結果を記憶したか否かを確認し (

S4015)、“Yes”の場合は処理を終了する。

【0113】

以上の処理により、パリティストレージ装置600の各パリティブロックには、追加されたデータストレージ装置600を含む、各データストレージ装置600のそれぞれ対応するパリティブロックのデータにより演算された排他的論理和の演算結果を記憶することができる。

【0114】

次に、図25の非同期方式パリティグループ設定画面において、「初期化方式」欄に“ランダム式”が入力される場合の処理の流れを図26に示す。

【0115】

まずパリティストレージ装置600は設定入力情報を受け付ける(S5000)。

そうするとパリティストレージ装置600は、まず「追加するデータストレージ装置」欄に記載されたデータストレージ装置600が最初のストレージ装置600であるか否かを判定する(S5001)。最初のストレージ装置600でない場合にはS5001において“No”に進む。

また最初のストレージ装置600である場合にはS5001において“Yes”に進む。そしてパリティストレージ装置600は、パリティブロック長を用いてパリティブロック数を求める(S5002)。

そしてパリティストレージ装置600は、パリティグループ管理テーブル710とパリティブロック管理テーブル720とを作成する(S5003)。

【0116】

続いてパリティストレージ装置600は、追加されるデータストレージ装置600へパリティブロック長を送信する(S5004)。

そうすると追加されるデータストレージ装置600は、パリティストレージ装置600から送信されたパリティブロック長を用いてパリティブロック数を求め(S5005)、パリティグループ管理テーブル710とパリティブロック管理テーブル720とを作成する(S5006)。

追加されるデータストレージ装置600は、パリティストレージ装置600へ、S5005にて求めたパリティブロック数を送信する(S5007)。

【0117】

パリティストレージ装置600は、追加されるデータストレージ装置600からパリティブロック数を受信すると、S5002で求めたパリティブロック数と比較し、パリティ生成可否を判定する(S5008)。S5002で求めたパリティブロック数よりも大きなパリティブロック数を追加されるデータストレージ装置600から受信した場合には、S5008において“否”に進み、データストレージ装置600の追加が不可である旨を管理端末160の出力装置166に表示して処理を終了する(S5010)。

【0118】

パリティストレージ装置600は、S5002で求めたパリティブロック数よりも大きなパリティブロック数を追加されるデータストレージ装置600から受信しない場合には、S5008において“可”に進み、パリティブロック数統括管理テーブル730を作成する(S5009)。以上で処理を終了する。

【0119】

その後、追加されたデータストレージ装置600は、情報処理装置200からハードディスクドライブへの書き込みデータを受信した場合に、書き込みデータと書き込みデータが書き込まれるパリティブロックに記憶されているデータとの排他的論理和を演算し、その演算結果を、書き込みデータが書き込まれるパリティブロックのパリティブロック番号と共にパリティストレージ装置600へ送信する。

【0120】

そして、パリティストレージ装置600は、追加されたデータストレージ装置600が情報処理装置200からハードディスクドライブへの書き込みデータを受信した場合に、

追加されたデータストレージ装置 6 0 0 により演算される、書き込みデータと書き込みデータが書き込まれるパリティブロックに記憶されているデータとの排他的論理和の演算結果を、書き込みデータが書き込まれるパリティブロックのパリティブロック番号と共に、追加されたデータストレージ装置 6 0 0 から受信し、受信した排他的論理和の演算結果と、受信したパリティブロック番号と同一のパリティブロック番号で特定されるパリティストレージ装置 6 0 0 のハードディスクドライブのパリティブロックに記憶されているデータとの排他的論理和を演算し、その排他的論理和の演算結果を、パリティストレージ装置 6 0 0 のパリティブロックに記憶する。

#### 【0 1 2 1】

このようにすることにより、パリティストレージ装置 6 0 0 の各パリティブロックには、追加されたデータストレージ装置 6 0 0 を含む各データストレージ装置 6 0 0 のそれぞれに対応するパリティブロックのデータにより演算された排他的論理和の演算結果を記憶することができる。

#### 【0 1 2 2】

次に、図 2 5 の非同期方式パリティグループ設定画面において、「削除するデータストレージ装置」欄にデータストレージ装置が入力される場合の処理の流れを図 3 2 に示す。

#### 【0 1 2 3】

データストレージ装置 6 0 0 をストレージシステムから削除する場合には、パリティストレージ装置 6 0 0 に記憶されているデータと、削除するデータストレージ装置 6 0 0 に記憶されているデータとの排他的論理和を演算し、その演算結果をパリティストレージ装置 6 0 0 に記憶するようにすればよい。

#### 【0 1 2 4】

これは、例えばデータストレージ装置 6 0 0 として、ストレージ装置 1 (6 0 0)、ストレージ装置 2 (6 0 0)、ストレージ装置 3 (6 0 0) を備えるストレージシステムからストレージ装置 3 (6 0 0) を削除する場合を考えると、ストレージ装置 1 (6 0 0) に記憶されるデータを D 1、ストレージ装置 2 (6 0 0) に記憶されるデータを D 2、ストレージ装置 3 (6 0 0) に記憶されるデータを D 3 とすれば、パリティストレージ装置 6 0 0 に記憶されるデータである D 4 は、 $D 1 (E X O R) D 2 (E X O R) D 3$  であるが、 $D 4 (E X O R) D 3$  を演算すると、 $D 1 (E X O R) D 2$  となるからである。

#### 【0 1 2 5】

データストレージ装置 6 0 0 を削除する場合の処理の流れを図 3 2 に示す。

まずパリティストレージ装置 6 0 0 は設定入力情報を受け付ける (S9000)。

#### 【0 1 2 6】

そうするとパリティストレージ装置 6 0 0 は、削除されるデータストレージ装置 6 0 0 へ、各パリティブロックに記憶されるデータをパリティブロック番号順に先頭から順次送信するように要求を送る (S9001)。そうすると削除されるデータストレージ装置 6 0 0 は、要求に従って、各パリティブロックに記憶されるデータの複製を先頭から順次、パリティブロック番号と共に送信する (S9002)。

#### 【0 1 2 7】

そしてパリティストレージ装置 6 0 0 は、削除されるデータストレージ装置 6 0 0 から、各パリティブロックのそれぞれに記憶されるデータの複製をパリティブロック番号と共に受信し、削除されるデータストレージ装置 6 0 0 から受信した各パリティブロックのデータの複製と、そのパリティブロック番号と同一のパリティブロック番号で特定されるパリティストレージ装置 6 0 0 のパリティブロックのデータ (旧パリティデータ) との排他的論理和を演算し (S9003)、その演算結果 (新パリティデータ) を旧パリティデータが記憶されていたパリティストレージ装置 6 0 0 のパリティブロックに記憶する (S9004)。そしてパリティストレージ装置 6 0 0 は、初期データ管理テーブル 7 4 0 を更新する (S9005)。以上の処理を全てのパリティブロックについて行う (S9006)。

#### 【0 1 2 8】

以上の処理により、パリティストレージ装置 6 0 0 の各パリティブロックには、削除さ

れるデータストレージ装置600を除く、各データストレージ装置600のそれぞれ対応するパリティブロックのデータにより演算された排他的論理和の演算結果を記憶することができる。

#### 【0129】

===バックアップデータの更新===

上記のようにして、パリティストレージ装置600が備えるハードディスクドライブの各パリティブロックに、各データストレージ装置600のデータをバックアップするためのデータ、すなわち各データストレージ装置600のそれぞれ対応するパリティブロックに記憶されるデータにより演算された排他的論理和の全ての演算結果、が記憶される。その後は、各データストレージ装置600が情報処理装置200からデータ書き込み要求を受信しハードディスクドライブに記憶されるデータが更新される毎に、パリティストレージ装置600に記憶されるデータストレージ装置600のデータをバックアップするためのデータの更新が行われる。パリティストレージ装置600に記憶されるデータストレージ装置600のデータをバックアップするためのデータを更新する際の処理の流れを図27に示す。

#### 【0130】

まずデータストレージ装置600は、情報処理装置200からデータ書き込み要求と共に書き込みデータ(DNEW)を受信する(S6000)。そうするとデータストレージ装置600は、その書き込みデータが書き込まれる記憶ボリューム310のパリティブロックに記憶されているデータ(旧データ、DOLD)を読み出す(S6001)。次にデータストレージ装置600は、書き込みデータと、その書き込みデータが書き込まれる記憶ボリューム310のパリティブロックに記憶されているデータとの排他的論理和を演算する(S6002)。そしてその排他的論理和の演算結果を、パリティブロック番号と共にパリティストレージ装置600に送信する(S6003)。

#### 【0131】

パリティストレージ装置600は、上記排他的論理和の演算結果をパリティブロック番号と共に受信すると、そのパリティブロック番号と同一のパリティブロック番号で特定されるパリティストレージ装置600のパリティブロックに記憶されているデータ(パリティデータ、POLD)を読み出す(S6004)。そして読み出したパリティデータと上記排他的論理和の演算結果との排他的論理和を演算する(S6005)。そして、その演算結果(PNEW)を、上記パリティブロックに記憶する(S6006)。

#### 【0132】

このようにすることにより、データストレージ装置600に記憶されるデータが情報処理装置200からのデータ書き込み要求により更新された場合に、パリティストレージ装置600に記憶される、データストレージ装置600のデータをバックアップするためのデータも更新することができる。

#### 【0133】

===データの復元===

次に、データストレージ装置600に記憶されるデータを復元する場合の処理の流れについて図28乃至31を用いて説明する。ここでは例として、データストレージ装置としてストレージ装置1(600)、ストレージ装置2(600)、ストレージ装置3(600)を備えるストレージシステムにおいて、ストレージ装置1(ある他のストレージ装置、第1のストレージ装置)(600)に記憶されるデータを復元する場合について説明する。また図29に示すように、ストレージ装置1(600)のパリティブロック数は5000000であり、ストレージ装置2(600)のパリティブロック数は6000000であり、ストレージ装置3(600)のパリティブロック数は7000000であり、パリティストレージ装置600のパリティブロック数は8000000であるとして説明する。

#### 【0134】

ここで、ストレージ装置1(600)に記憶されているデータを復元するやり方として2つの方式が考えられる。すなわち、第1の方式は、パリティストレージ装置600が、

ストレージ装置 1 (600) のデータを復元するために必要な 500000 個のパリティブロックのデータをストレージ装置 2 (600) 及びストレージ装置 3 (600) から受信し、それによりストレージ装置 1 (600) のデータを復元するやり方である。一方、第 2 の方式は、パリティストレージ装置 600 が、パリティストレージ装置 600 のパリティブロック数である 800000 個のパリティブロックのデータをストレージ装置 2 (600) 及びストレージ装置 3 (600) から受信し、それにより 800000 個のパリティブロックのデータを復元してストレージ装置 1 (600) に送信するやり方である。第 1 の方式でデータの復元を行う場合には、データの復元時にストレージ装置 600 間で授受されるデータの転送量が最小限で済むという利点がある。またそのためデータの回復を比較的短時間で行うことが可能であるという利点がある。一方第 2 の方式でデータの復元を行う場合には、パリティストレージ装置 600 は、各データストレージ装置 600 のパリティブロック数を管理する必要が無い、すなわち図 12 に示したパリティブロック数統括管理テーブル 730 を共有メモリ 120 に記憶しておく必要がないという利点がある。

#### 【0135】

まず第 1 の方式を図 28、図 29 を用いて説明する。

ストレージ装置 1 (600) は、障害検知プログラム 840 によりハードディスクドライブに記憶されているデータに異常を検出すると (S7000)、パリティストレージ装置 600 に対して、ハードディスクドライブの各パリティブロックに記憶されているべき元のデータの送信要求を送信する (S7001)。

#### 【0136】

パリティストレージ装置 600 は、上記送信要求を受信すると、その送信要求に応じて、パリティブロック数統括管理テーブル 730 を参照し、ストレージ装置 1 (600) 以外のストレージ装置 (第 1 の記憶データの送信要求を送信した第 1 のストレージ装置以外の第 1 のストレージ装置) 600、すなわち、ストレージ装置 2 (600) とストレージ装置 3 (600) とに、それぞれ、パリティブロック番号 0～499999 の合計 500000 個のパリティブロックのデータの複製を、各パリティブロック番号と共に送信させるための送信要求を送信する (S7002)。

#### 【0137】

そうすると、ストレージ装置 2 (600) 及びストレージ装置 3 (600) のそれぞれは、上記送信要求に従い、パリティブロック番号 0～499999 のパリティブロックのデータの複製を、各パリティブロック番号と共に送信する (S7003、S7004)。

#### 【0138】

パリティストレージ装置 600 は、ストレージ装置 2 (600) 及びストレージ装置 3 (600) のそれぞれから、パリティブロック番号 0～499999 のパリティブロックのデータの複製を、各パリティブロック番号と共に受信すると、パリティブロック番号が同一の上記受信したデータの複製と、そのパリティブロック番号と同一のパリティストレージ装置 600 のパリティブロックに記憶されているデータとの排他的論理和を演算する (S7005)。そしてその排他的論理和の演算結果を、パリティブロック番号と共に、ストレージ装置 1 (600) に送信する (S7006、S7007)。

そしてストレージ装置 1 (600) は、その演算結果をパリティブロック番号と共に受信し、そのパリティブロック番号で特定される記憶ボリューム 310 のパリティブロックに書き込む (S7008、S7009)。

これにより、ストレージ装置 1 (600) に記憶されるデータを復元することができる。

#### 【0139】

次に第 2 の方式を図 30、図 31 を用いて説明する。

ストレージ装置 1 (600) は、障害検知プログラム 840 によりハードディスクドライブに記憶されているデータに異常を検出すると (S8000)、パリティストレージ装置 600 に対してハードディスクドライブの各パリティブロックに記憶されているべき元のデ

ータの送信要求を送信する (S8001)。

【0140】

パリティストレージ装置 6 0 0 は、上記送信要求を受信すると、その送信要求に応じて、ストレージ装置 1 (6 0 0) 以外のストレージ装置 (第 1 の記憶データの送信要求を送信した第 1 のストレージ装置以外の第 1 のストレージ装置) 6 0 0、すなわち、ストレージ装置 2 (6 0 0) とストレージ装置 3 (6 0 0) とに、それぞれ、パリティブロック番号 0 ~ 7 9 9 9 9 9 の合計 8 0 0 0 0 0 個のパリティブロックのデータの複製を、各パリティブロック番号と共に送信させるための送信要求を送信する (S8002)。

【0141】

そうすると、ストレージ装置 2 (6 0 0) 及びストレージ装置 3 (6 0 0) のそれぞれは、上記送信要求に従い、パリティブロック番号 0 ~ 7 9 9 9 9 9 のパリティブロックのデータの複製を、各パリティブロック番号と共に送信する (S8003、S8004)。ここで、ストレージ装置 2 (6 0 0) についてはパリティブロック数が 6 0 0 0 0 0 であるので、パリティブロック番号 6 0 0 0 0 0 から 7 9 9 9 9 9 のパリティブロックは存在しない。そのためストレージ装置 2 (6 0 0) は、パリティブロック番号 6 0 0 0 0 0 から 7 9 9 9 9 9 のパリティブロックについては、0 が記憶されているものとして、パリティストレージ装置 6 0 0 にデータを送信するようにする。同様に、ストレージ装置 3 (6 0 0) についてはパリティブロック数が 7 0 0 0 0 0 であるので、パリティブロック番号 7 0 0 0 0 0 から 7 9 9 9 9 9 のパリティブロックは存在しない。そのためストレージ装置 3 (6 0 0) は、パリティブロック番号 7 0 0 0 0 0 から 7 9 9 9 9 9 のパリティブロックについては、0 が記憶されているものとして、パリティストレージ装置 6 0 0 にデータを送信するようにする。

【0142】

パリティストレージ装置 6 0 0 は、ストレージ装置 2 (6 0 0) 及びストレージ装置 3 (6 0 0) のそれぞれから、パリティブロック番号 0 ~ 7 9 9 9 9 9 のパリティブロックのデータの複製を、各パリティブロック番号と共に受信すると、パリティブロック番号が同一の上記受信したデータの複製と、そのパリティブロック番号と同一のパリティストレージ装置 6 0 0 のパリティブロックに記憶されているデータとの排他的論理和を演算する (S8005)。そしてその排他的論理和の演算結果を、パリティブロック番号と共に、ストレージ装置 1 (6 0 0) に送信する (S8006、S8007)。

【0143】

そしてストレージ装置 1 (6 0 0) は、その演算結果をパリティブロック番号と共に受信し、そのパリティブロック番号で特定される記憶ボリューム 3 1 0 のパリティブロックに書き込む (S8008、S8009)。ここで、ストレージ装置 1 (6 0 0) についてはパリティブロック数が 5 0 0 0 0 0 であるので、パリティブロック番号 5 0 0 0 0 0 から 7 9 9 9 9 9 のパリティブロックは存在しない。そのためストレージ装置 1 (6 0 0) は、パリティブロック番号 5 0 0 0 0 0 から 7 9 9 9 9 9 のパリティブロックについては、記憶ボリューム 3 1 0 へ書き込まないように制御する。

このように第 2 の方式によっても、ストレージ装置 1 (6 0 0) に記憶されるデータを復元することができる。

【0144】

なお上述したデータストレージ装置 6 0 0 に記憶されるデータを復元する場合の処理は、既にパリティストレージ装置 6 0 0 のハードディスクドライブの各パリティブロックには、全てのデータストレージ装置 6 0 0 のパリティブロックに記憶されるデータから演算された排他的論理和の全ての演算結果が記憶されている場合についてのものである。

【0145】

しかし、ストレージシステムにデータストレージ装置 6 0 0 が追加され、パリティストレージ装置 6 0 0 のパリティブロックに記憶されるデータを非同期方式で更新している間に、あるデータストレージ装置 6 0 0 から、そのデータストレージ装置 6 0 0 のハードディスクドライブに記憶されていた元のデータの復元が要求される場合がある。

この場合、パリティストレージ装置 600 のどのパリティブロックに記憶されるデータが更新済みであるかを、図 13 (B) 又は図 13 (C) に示す初期データ管理テーブル 740 に記憶するようにしている場合には、データストレージ装置 600 の追加に伴うパリティストレージ装置 600 のパリティブロックに記憶されるデータの更新が終了するのを待たずに、データの復元を開始することが可能である。

#### 【0146】

すなわち、パリティストレージ装置 600 が、障害を検知したデータストレージ装置 600 以外のデータストレージ装置 600 に、それぞれの各パリティブロックのデータの複製を各パリティブロック番号と共に送信させるための送信要求を送信する際に、更新済みのパリティブロックのパリティブロック番号で特定されるパリティブロックのデータについては、追加されたデータストレージ装置 600 を含む、各データストレージ装置 600 に送信要求を送信するようにし、未更新のパリティブロックのパリティブロック番号で特定されるパリティブロックのデータについては、追加されたデータストレージ装置 600 を含まない、各データストレージ装置 600 に送信要求を送信するようにする。

このようにすれば、障害を検知したデータストレージ装置 600 に対して、いち早く復元したデータを送信することが可能となる。

#### 【0147】

以上、本実施の形態に係るストレージ装置 600 について説明したが、本実施の形態によれば、パリティストレージ装置 600 の各パリティブロックには、複数のデータストレージ装置 600 のそれぞれ対応する各パリティブロックのデータにより演算された排他的論理和の演算結果を記憶することができる。これにより、データストレージ装置 600 の記憶領域に記憶されるデータが例えば消失し、そのデータを復元することが必要となった場合には、そのデータストレージ装置以外のストレージ装置 600 の記憶領域に記憶されるデータを用いて排他的論理和を演算することにより、消失したデータを復元することができる。

#### 【0148】

つまり本実施の形態によれば、パリティストレージ装置 600 と通信可能に接続されるデータストレージ装置 600 の台数が何台であっても、データストレージ装置 600 の記憶領域に記憶されるデータをバックアップすることが可能となる。これにより、データのバックアップを合理化し、データのバックアップを行うために必要とされるストレージ装置 600 の数を抑制することができる。このためストレージ装置 600 を維持管理するためのコスト、さらには情報処理システムを維持管理するためのコストも抑制することも可能となる。

#### 【0149】

また本実施の形態に係るストレージ装置 600 は、パリティストレージ装置 600 とデータストレージ装置 600 とで、それぞれの記憶領域をパリティデータ長毎に論理的に区分するようにしている。これにより、両ストレージ装置 600 の記憶領域をパリティブロック毎に対応付けることが可能となる。つまり、パリティストレージ装置 600 とデータストレージ装置 600 とで同一のパリティブロック番号で特定されるパリティブロック同士を対応づけることが可能となる。このようにすることにより、メインフレーム系情報処理システムで使用されることの多い CKD 方式や、オープン系情報処理システムで使用されることの多い FBA 方式等の、記憶ボリューム 310 の制御の方式が異なるストレージ装置 600 を用いてストレージシステムを構成する場合にも、データストレージ装置 600 に記憶されるデータをバックアップすることができるようになる。

#### 【0150】

以上発明を実施するための最良の形態について説明したが、上記実施の形態は本発明の理解を容易にするためのものであり、本発明を限定して解釈するためのものではない。本発明はその趣旨を逸脱することなく変更、改良され得ると共に、本発明にはその等価物も含まれる。

#### 【図面の簡単な説明】



## 【0151】

- 【図1】 本実施の形態に係る情報処理システムの全体構成を示すブロック図である。
- 【図2】 本実施の形態に係るストレージ装置の全体構成を示すブロック図である。
- 【図3】 本実施の形態に係るストレージ装置の外観構成を示す図である。
- 【図4】 本実施の形態に係るストレージ制御装置の外観構成を示す図である。
- 【図5】 本実施の形態に係る管理端末の構成を示すブロック図である。
- 【図6】 本実施の形態に係るチャンネル制御部の構成を示すブロック図である。
- 【図7】 本実施の形態に係るディスク制御部の構成を示すブロック図である。
- 【図8】 本実施の形態に係る情報処理装置の構成を示すブロック図である。
- 【図9】 本実施の形態に係る共有メモリに記憶されるテーブルを示す図である。
- 【図10】 本実施の形態に係るパリティ管理テーブルを示す図である。
- 【図11】 本実施の形態に係るパリティブロック管理テーブルを示す図である。
- 【図12】 本実施の形態に係るパリティブロック数統括管理テーブルを示す図である。

。 【図13】 本実施の形態に係る初期データ管理テーブルを示す図である。

【図14】 本実施の形態に係るストレージ装置における処理の流れを示すフローチャートである。

【図15】 本実施の形態に係る設定画面例を示す図である。

【図16】 本実施の形態に係るストレージ装置における処理の内容を示す図である。

【図17】 本実施の形態に係るストレージ装置における処理の内容を示す図である。

【図18】 本実施の形態に係るストレージ装置における処理の内容を示す図である。

【図19】 本実施の形態に係るストレージ装置における処理の内容を示す図である。

【図20】 本実施の形態に係るストレージ装置における処理の内容を示す図である。

【図21】 本実施の形態に係るストレージ装置における処理の流れを示すフローチャートである。

【図22】 本実施の形態に係るストレージ装置における処理の内容を示す図である。

【図23】 本実施の形態に係るストレージ装置における処理の流れを示すフローチャートである。

【図24】 本実施の形態に係るストレージ装置における処理の流れを示すフローチャートである。

【図25】 本実施の形態に係る設定画面例を示す図である。

【図26】 本実施の形態に係るストレージ装置における処理の流れを示すフローチャートである。

【図27】 本実施の形態に係るストレージ装置における処理の流れを示すフローチャートである。

【図28】 本実施の形態に係るストレージ装置における処理の流れを示すフローチャートである。

【図29】 本実施の形態に係るストレージ装置における処理の内容を示す図である。

【図30】 本実施の形態に係るストレージ装置における処理の流れを示すフローチャートである。

【図31】 本実施の形態に係るストレージ装置における処理の内容を示す図である。

【図32】 本実施の形態に係るストレージ装置における処理の流れを示すフローチャートである。

【図33】 本実施の形態に係るストレージ装置間でのデータ転送フォーマットの一例を示す図である。

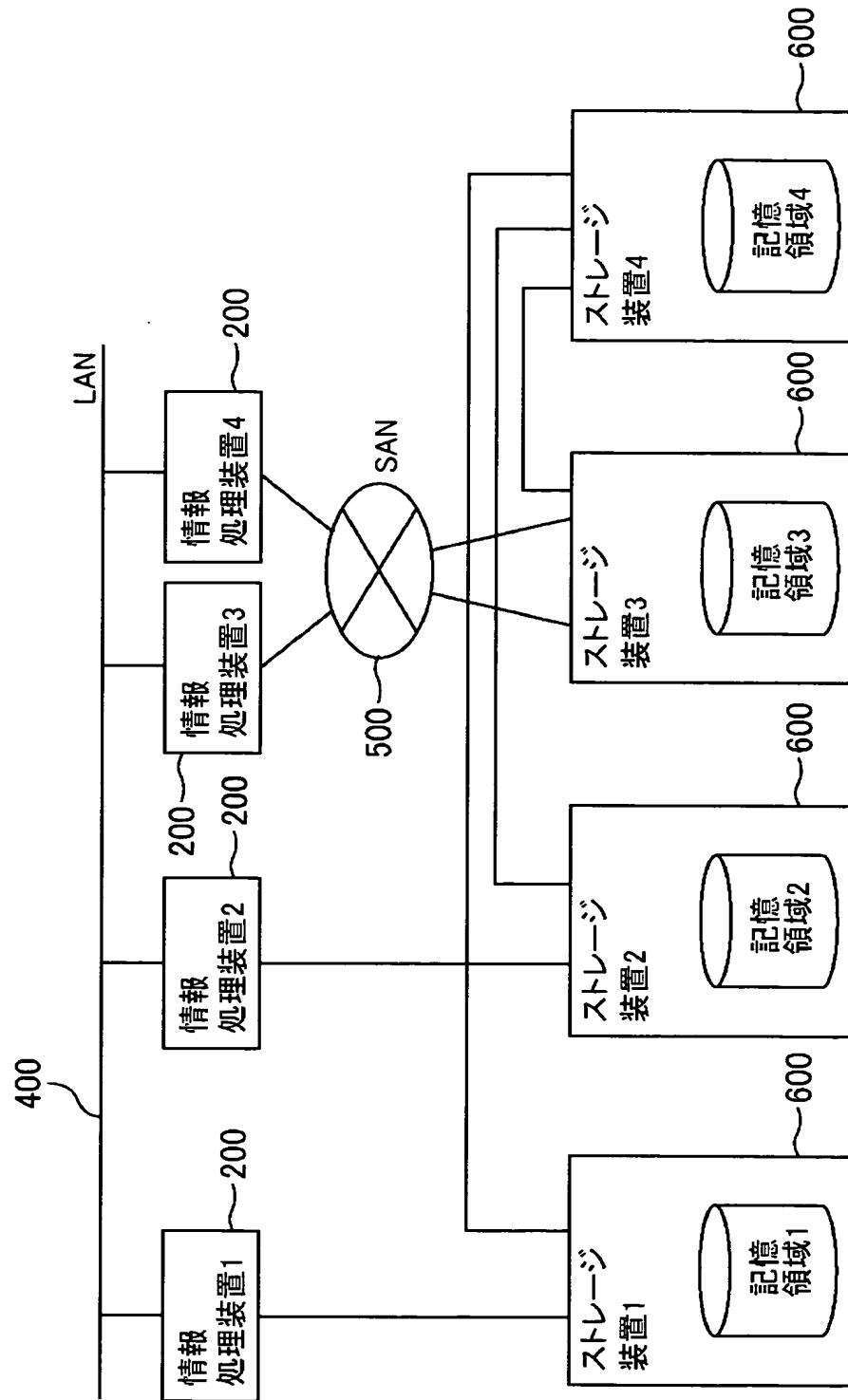
## 【符号の説明】

## 【0152】

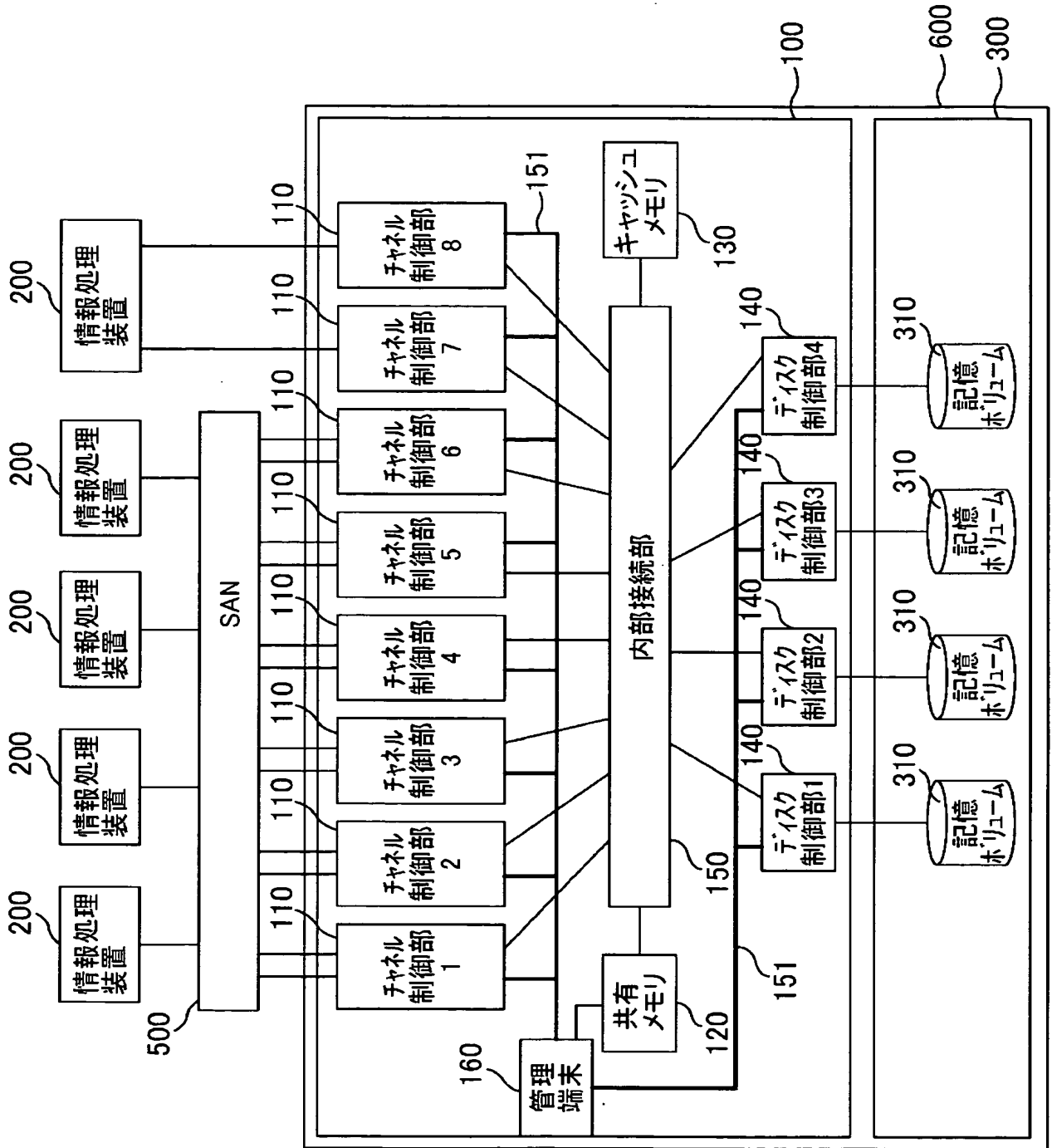
- 100      ストレージ制御装置
- 110      チャンネル制御部
- 120      共有メモリ

1 3 0	キャッシュメモリ
1 4 0	ディスク制御部
1 5 0	接続部
1 6 0	管理端末
2 0 0	情報処理装置
3 0 0	ストレージ駆動装置
3 1 0	記憶ボリューム
5 0 0	S A N
6 0 0	ストレージ装置
7 1 0	パリティグループ管理テーブル
7 2 0	パリティブロック管理テーブル
7 3 0	パリティブロック数統括管理テーブル
7 4 0	初期データ管理テーブル
8 1 0	ストレージ制御プログラム
8 2 0	データ送受信制御プログラム
8 3 0	パリティ制御プログラム
8 4 0	障害検出プログラム

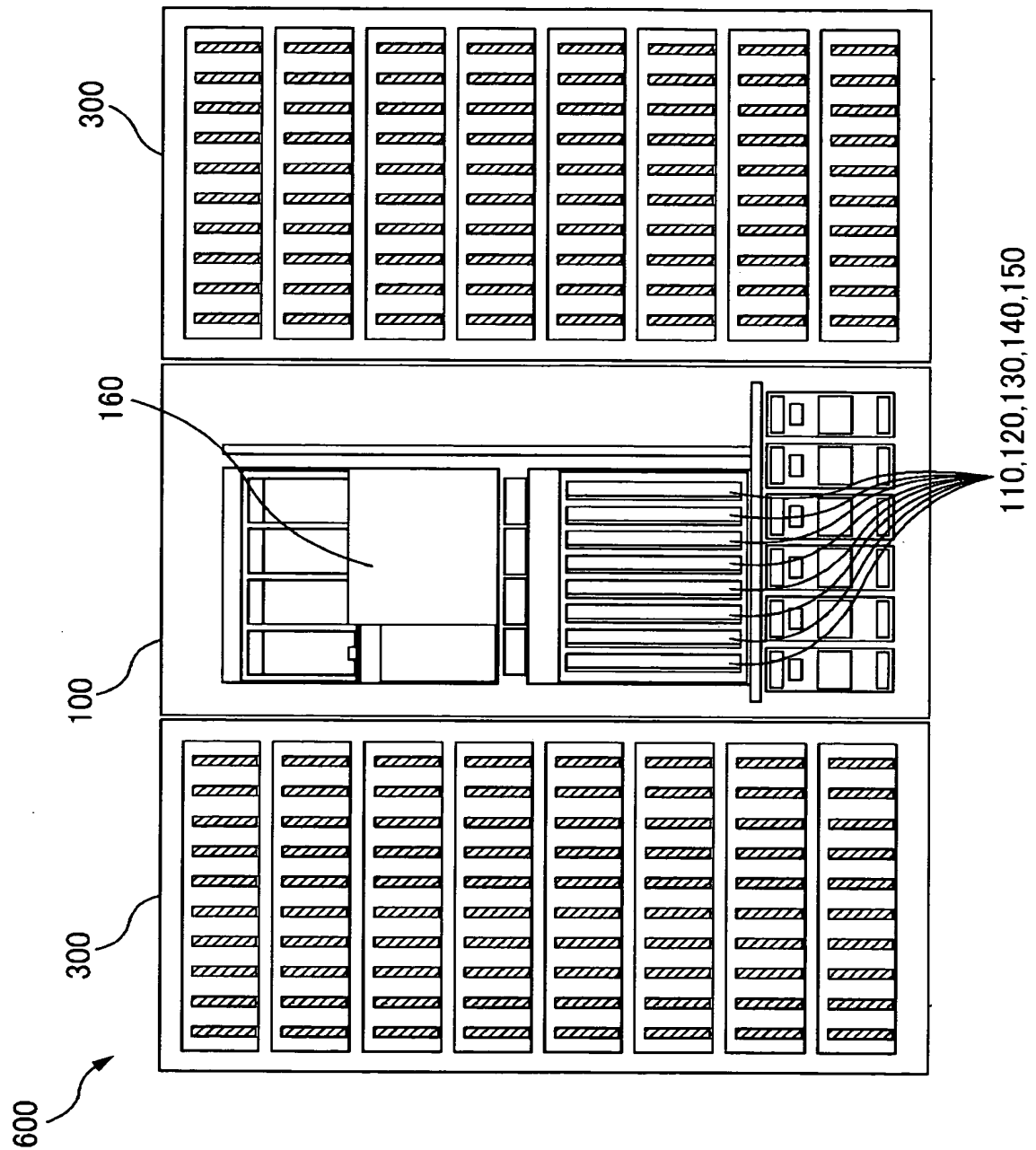
【書類名】 図面  
【図 1】



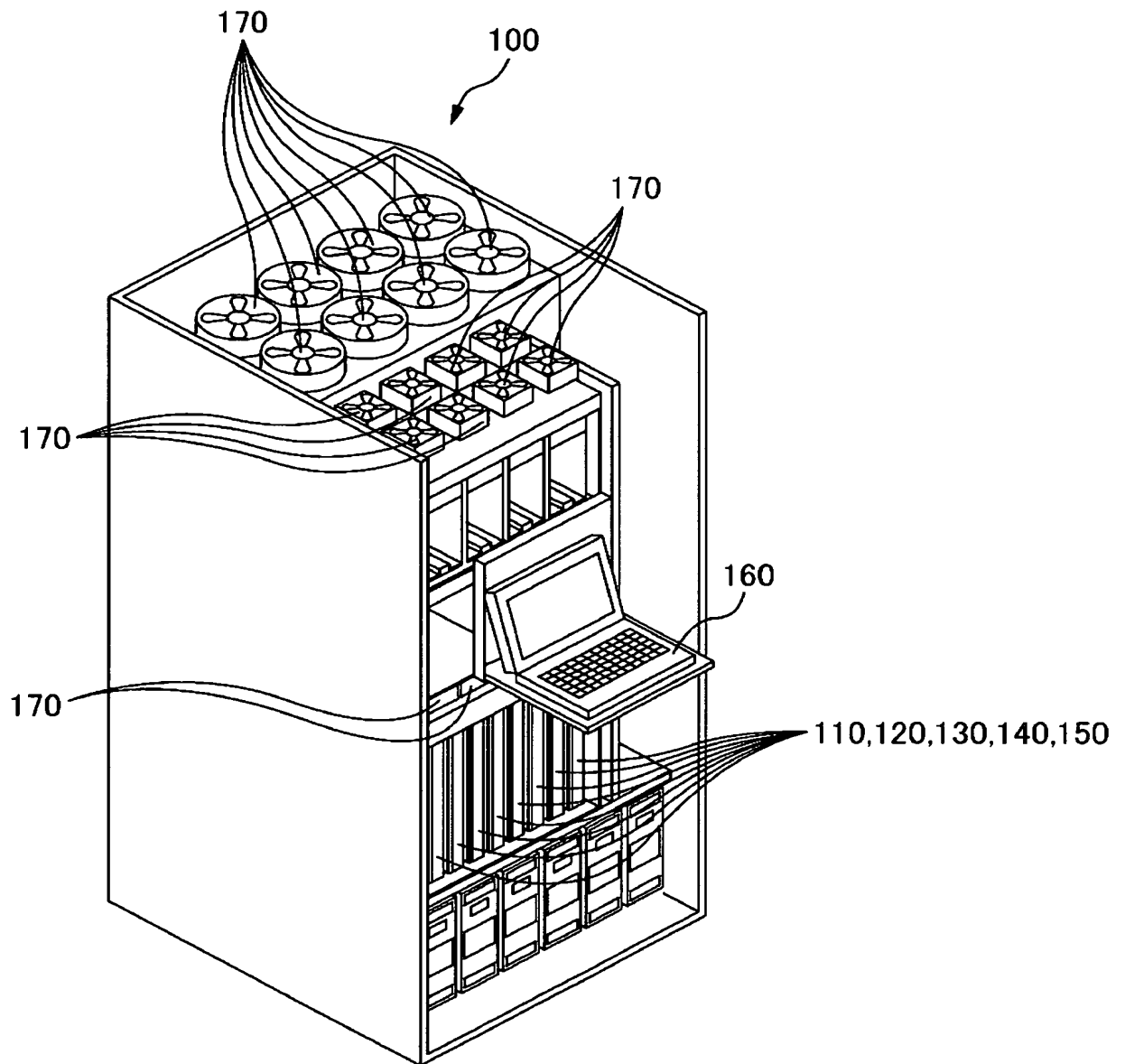
【図 2】



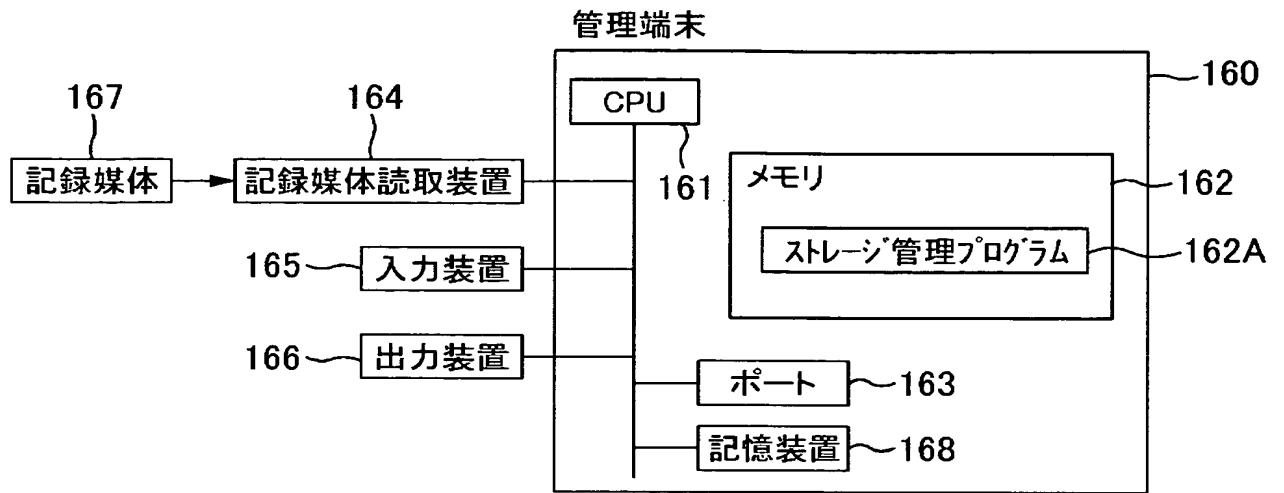
【図 3】



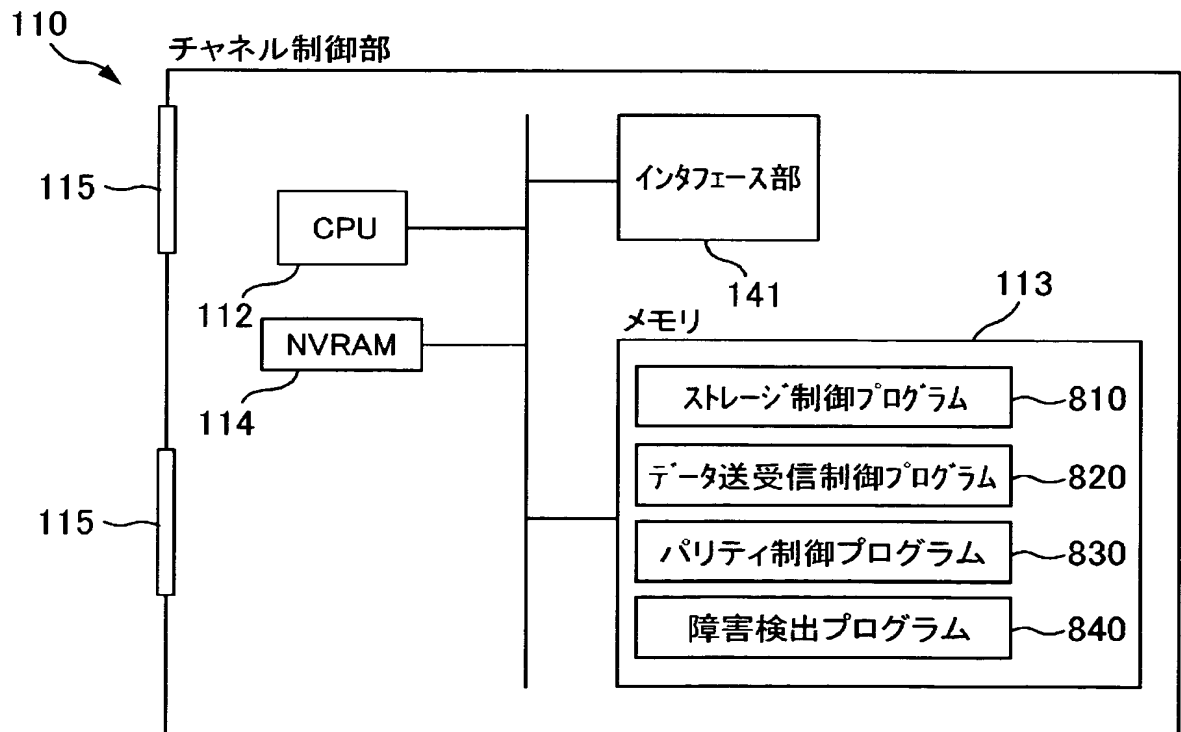
【図 4】



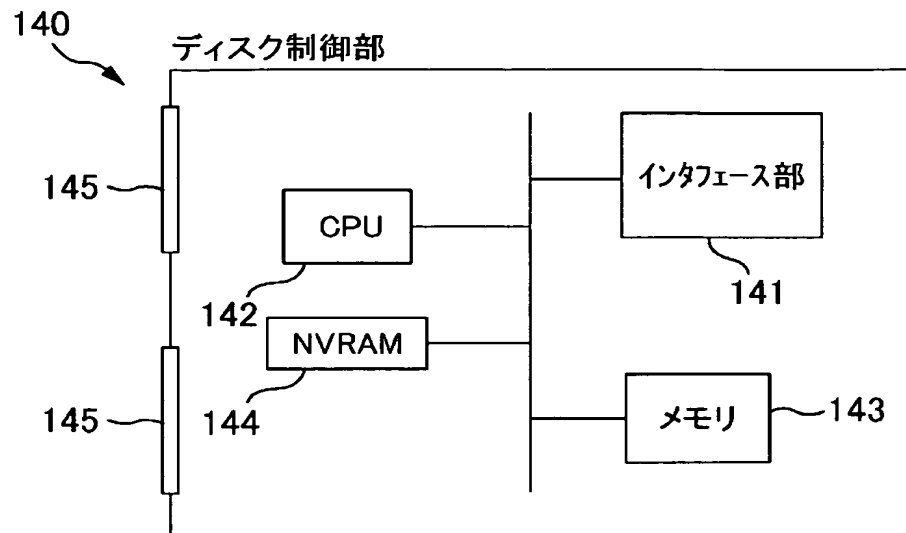
【図 5】



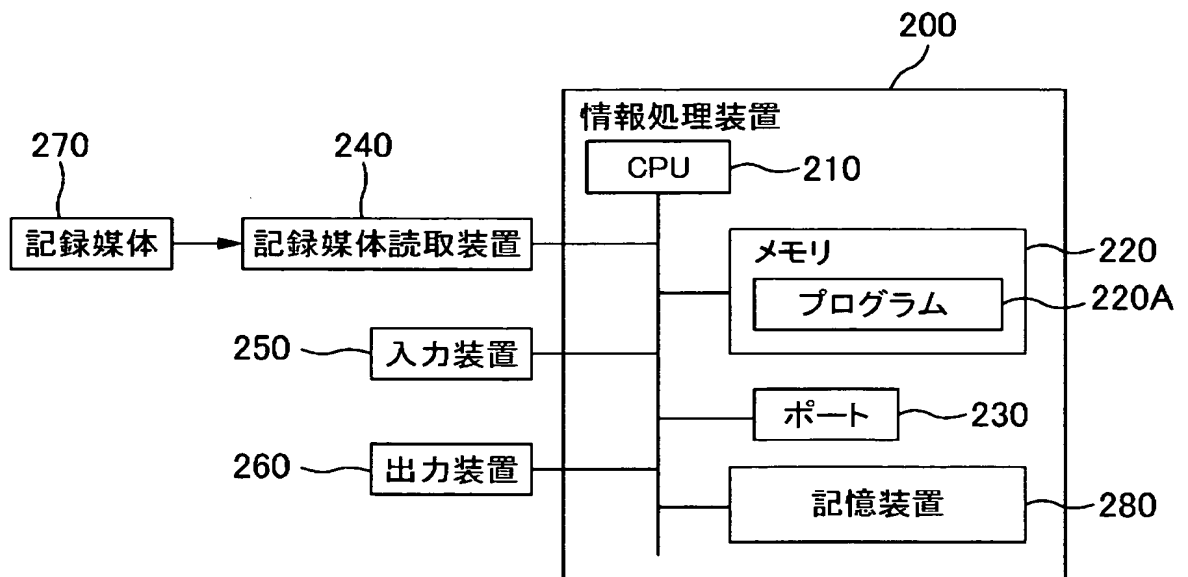
【図 6】



【図 7】

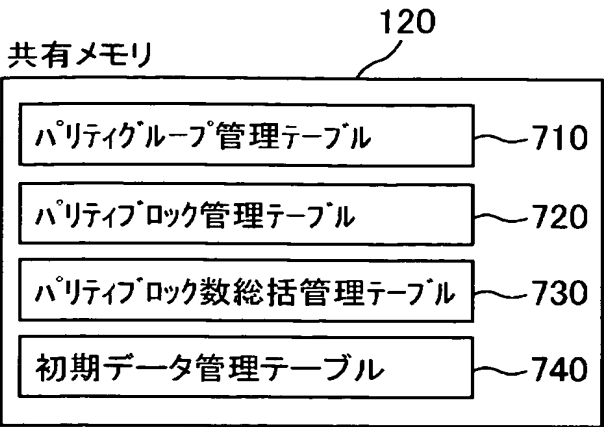


【図 8】





【図 9】



【図 1 0】

パリティグループ管理テーブル 710

データ記憶 ストレージ装置	ストレージ装置1,ストレージ装置2,ストレージ装置3
パリティ記憶 ストレージ装置	ストレージ装置4

【図 11】

パリティブロック管理テーブル 720

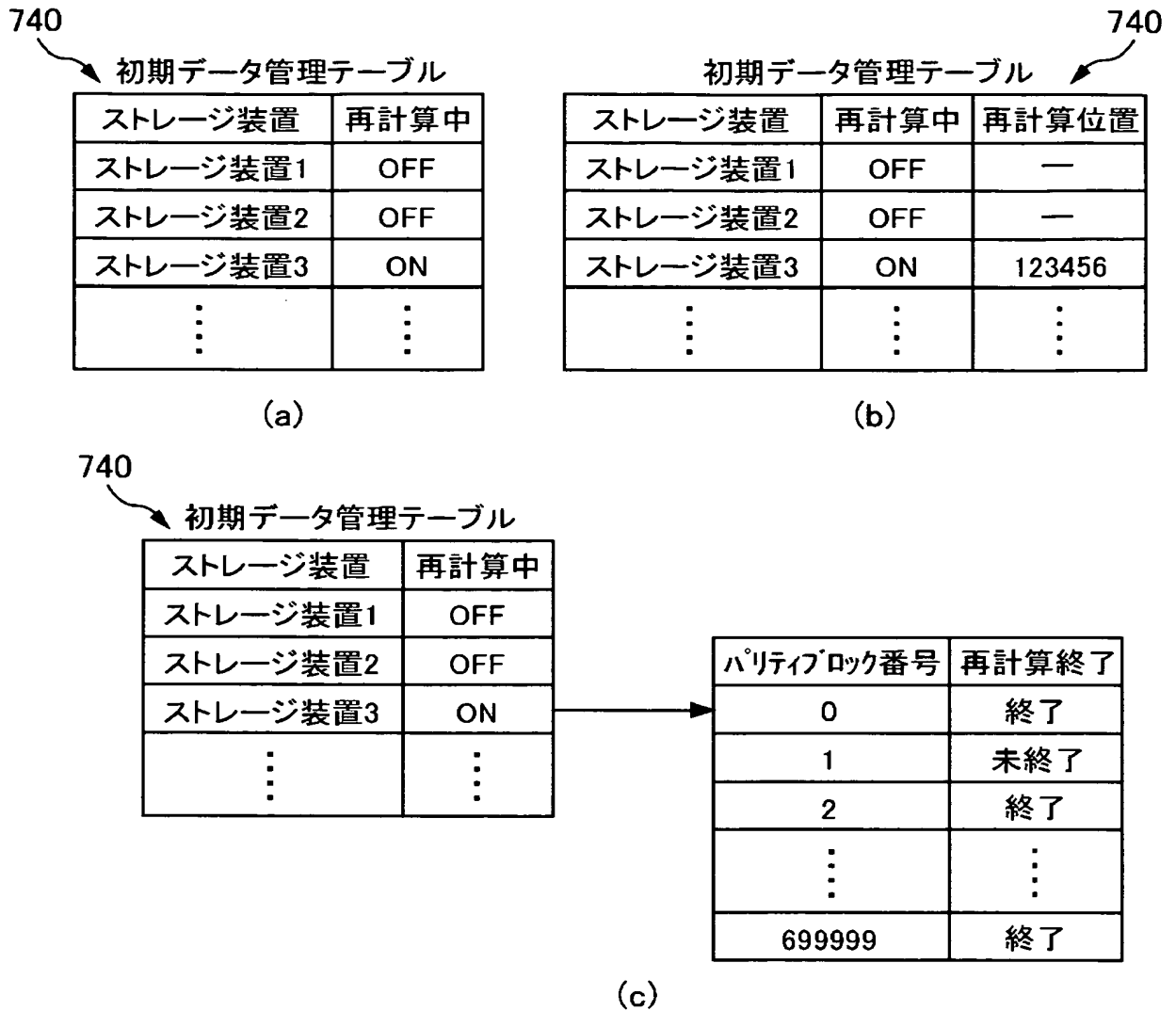
パリティブロック長=512バイト	
パリティブロック数=800000	
論理ボリューム番号	先頭パリティブロック番号
0000	0
0001	4096
0002	16384
0003	40000
0004	81920
⋮	⋮

【図 12】

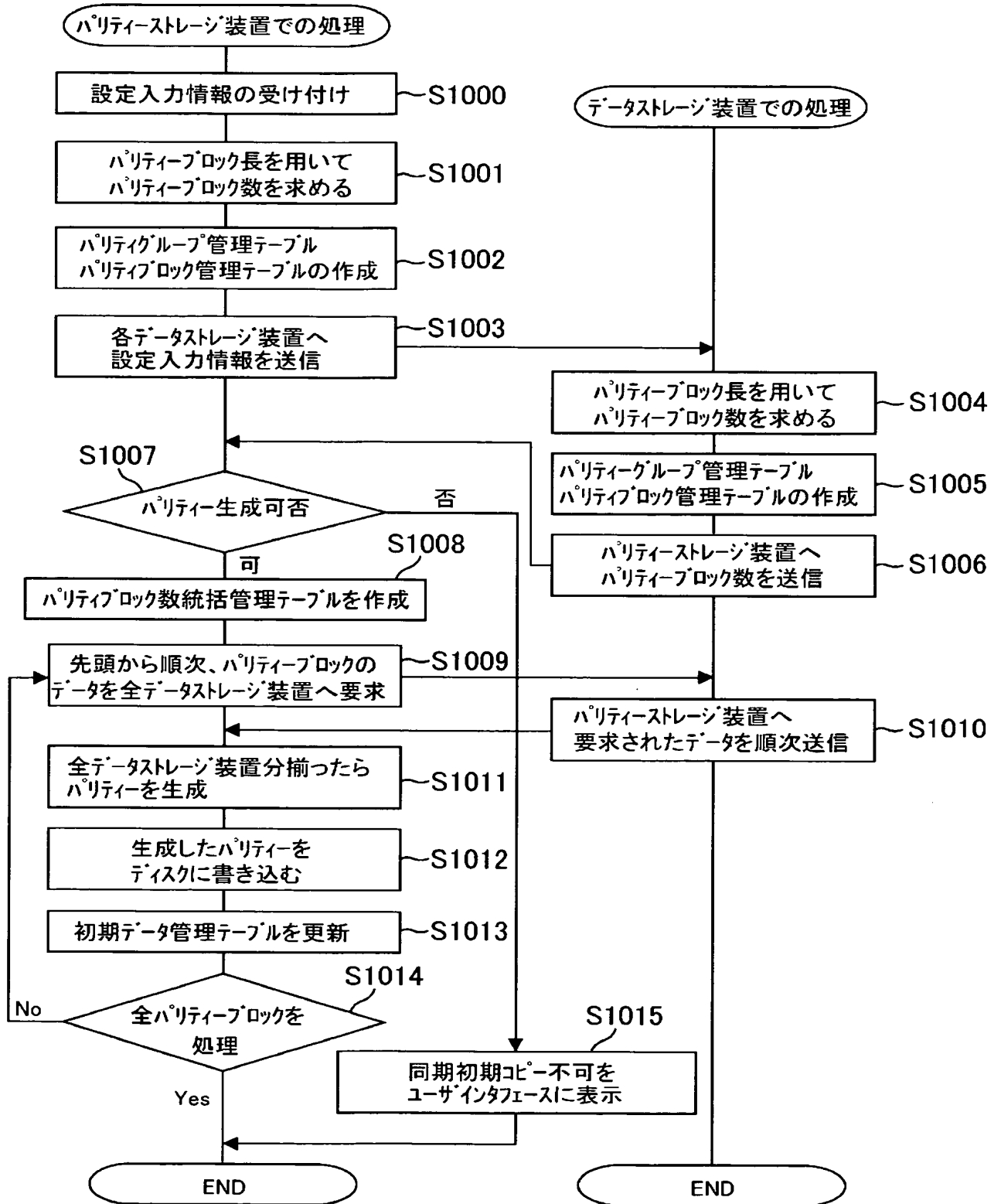
パリティブロック数統括管理テーブル 730

ストレージ装置	パリティブロック数
ストレージ装置1	500000
ストレージ装置2	600000
ストレージ装置3	700000
ストレージ装置4	800000

【図 13】



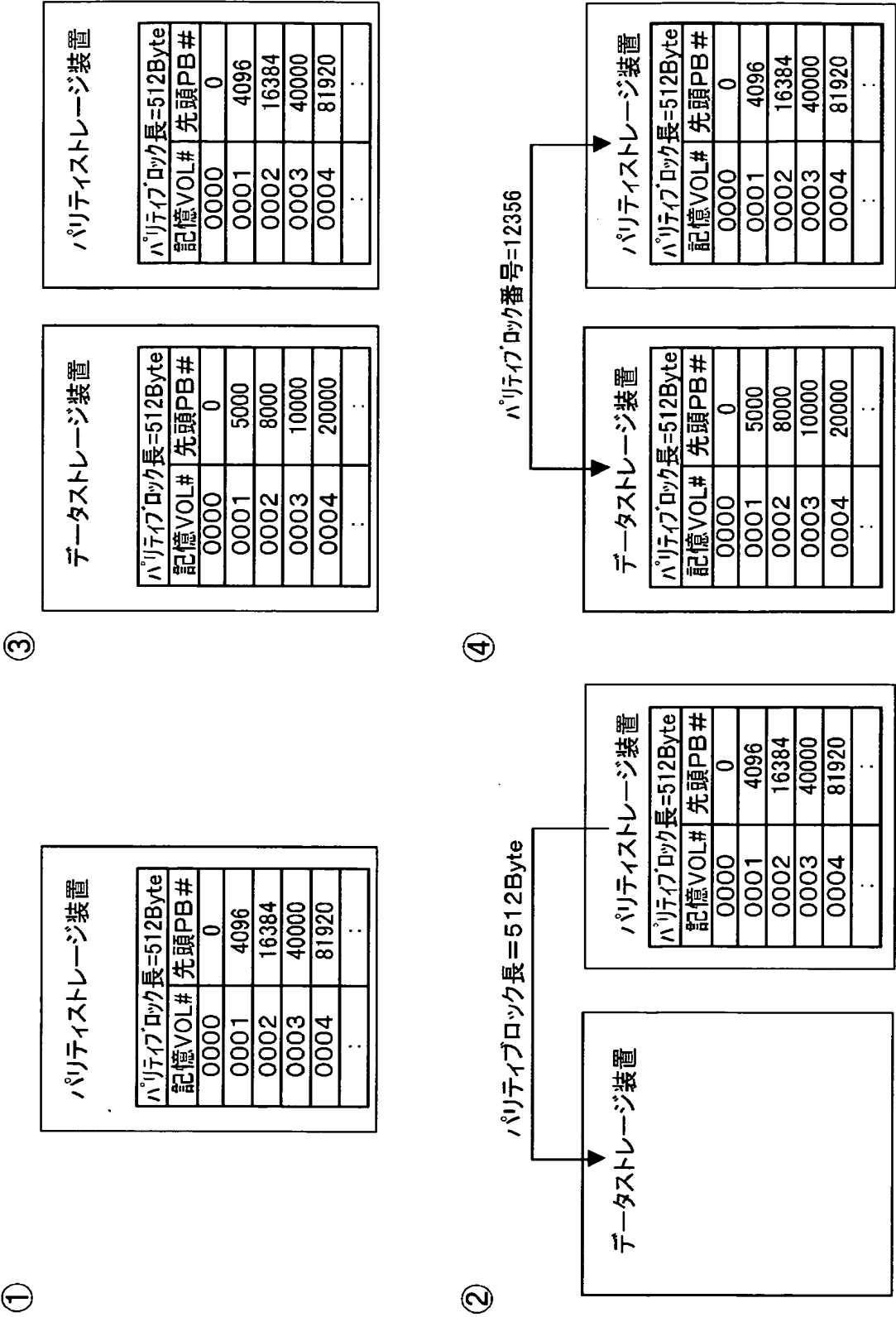
【図 14】



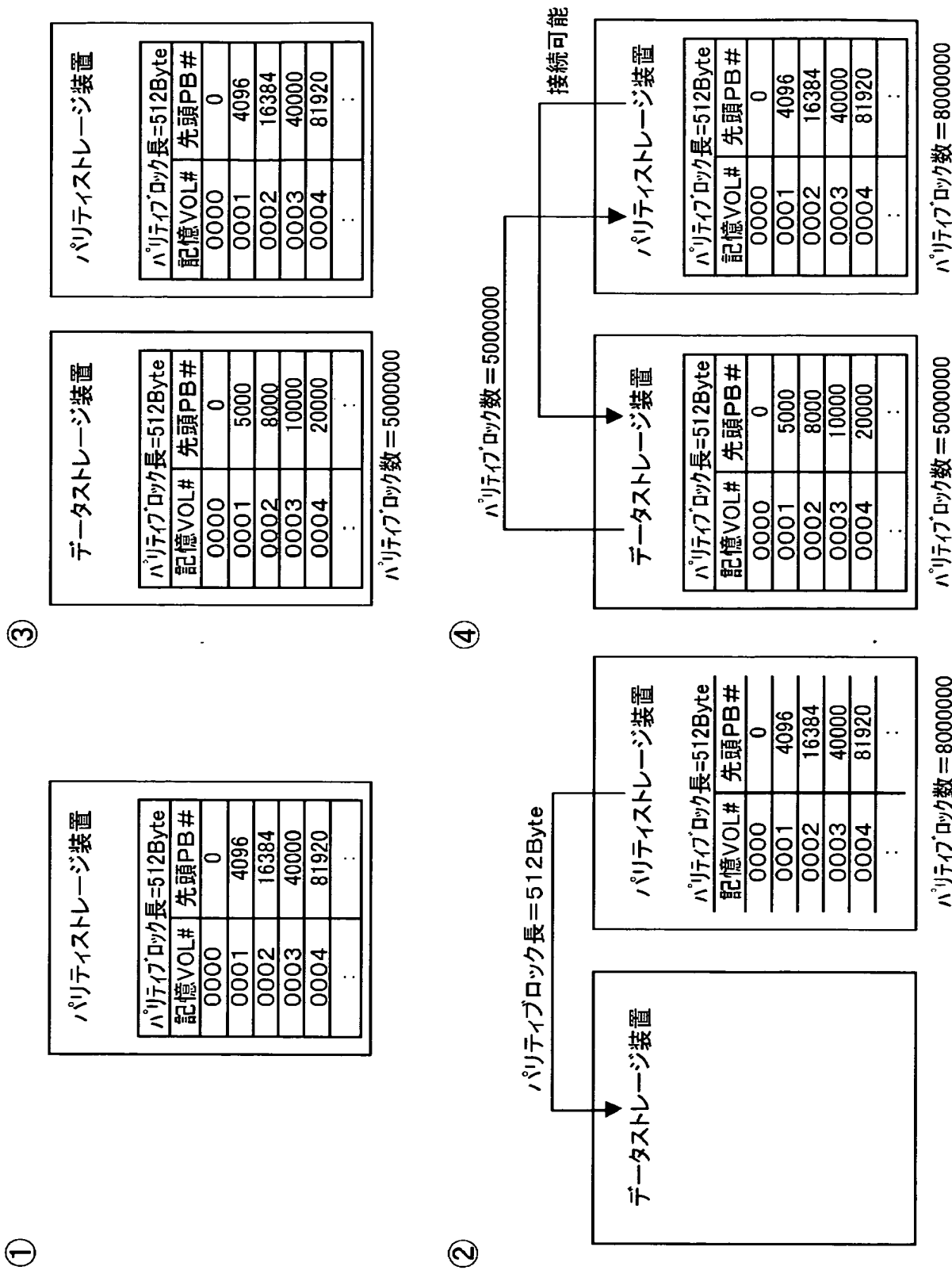
【図 15】

同期方式パリティグループ設定画面		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
パリティブロック長	<input type="text" value="512"/>	バイト		
パリティグループ 構成ストレージ装置	<div><div>ストレージ装置1</div><div>ストレージ装置2</div><div>ストレージ装置3</div></div> <div>▲ ▼</div>			
パリティデータ 記憶ストレージ装置	<input type="text" value="ストレージ装置4"/>			
初期化方式	<input type="text" value="順次式"/>			
		<input type="button" value="OK"/>	<input type="button" value="Cancel"/>	

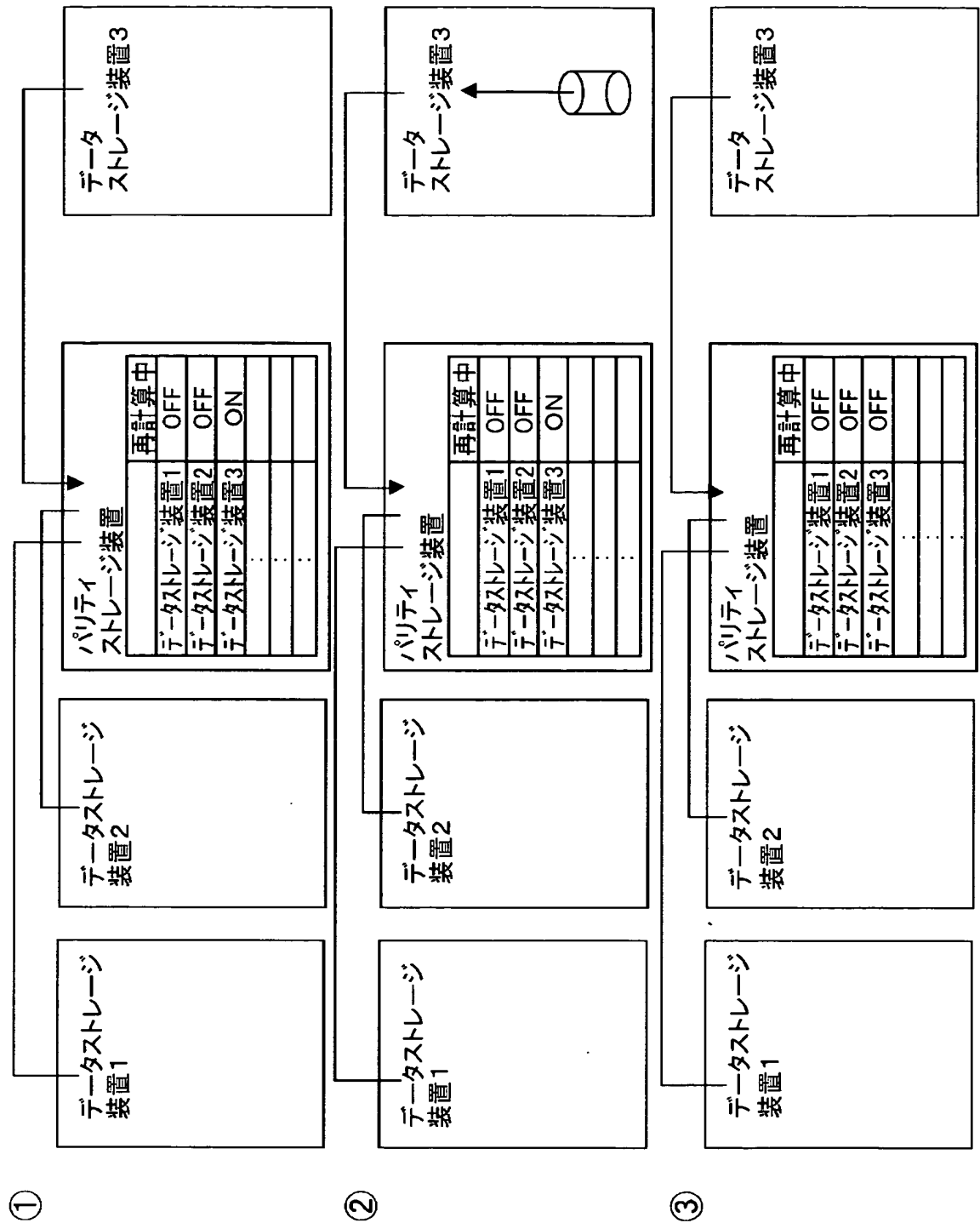
【図 16】



【図 17】

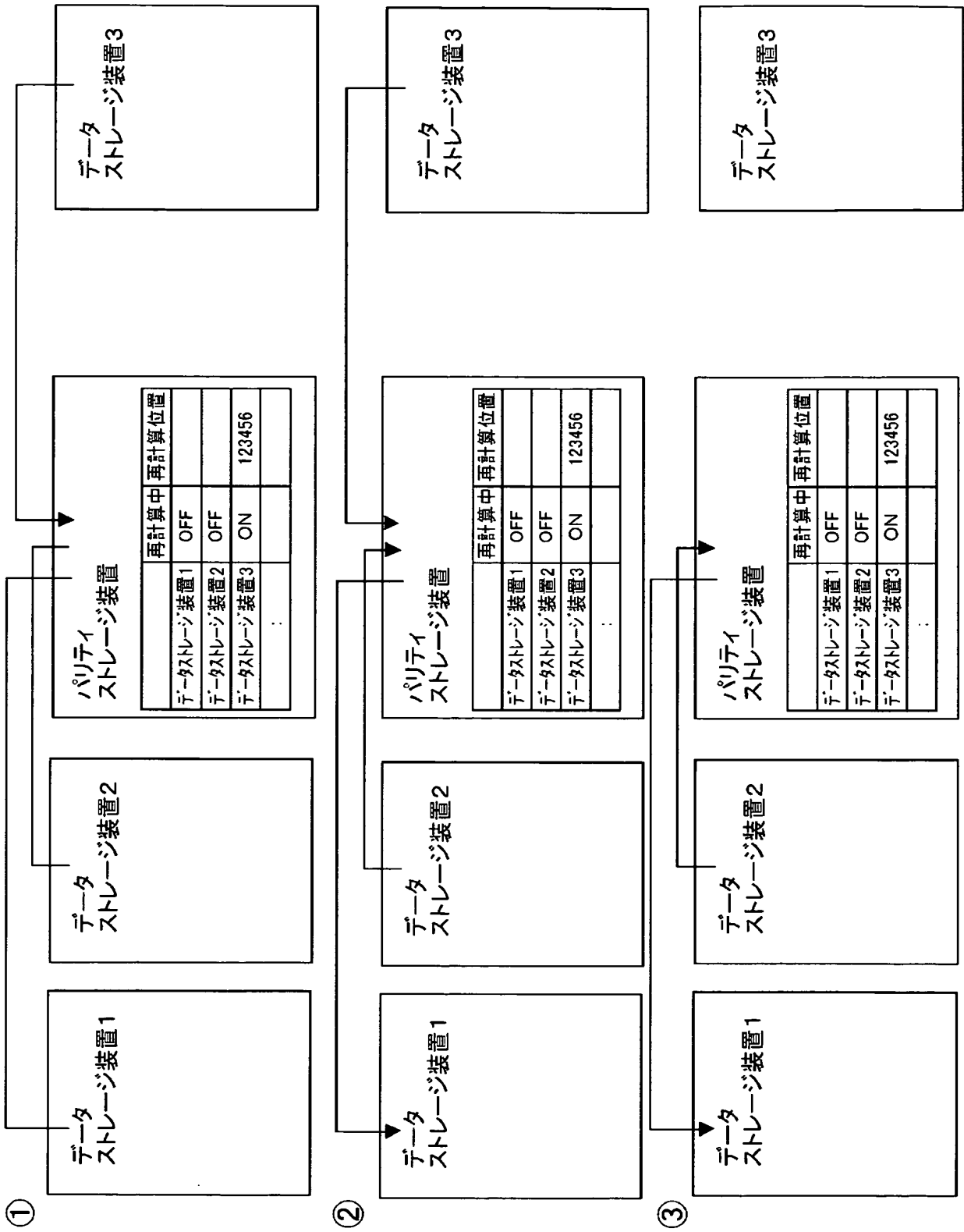


【図 18】

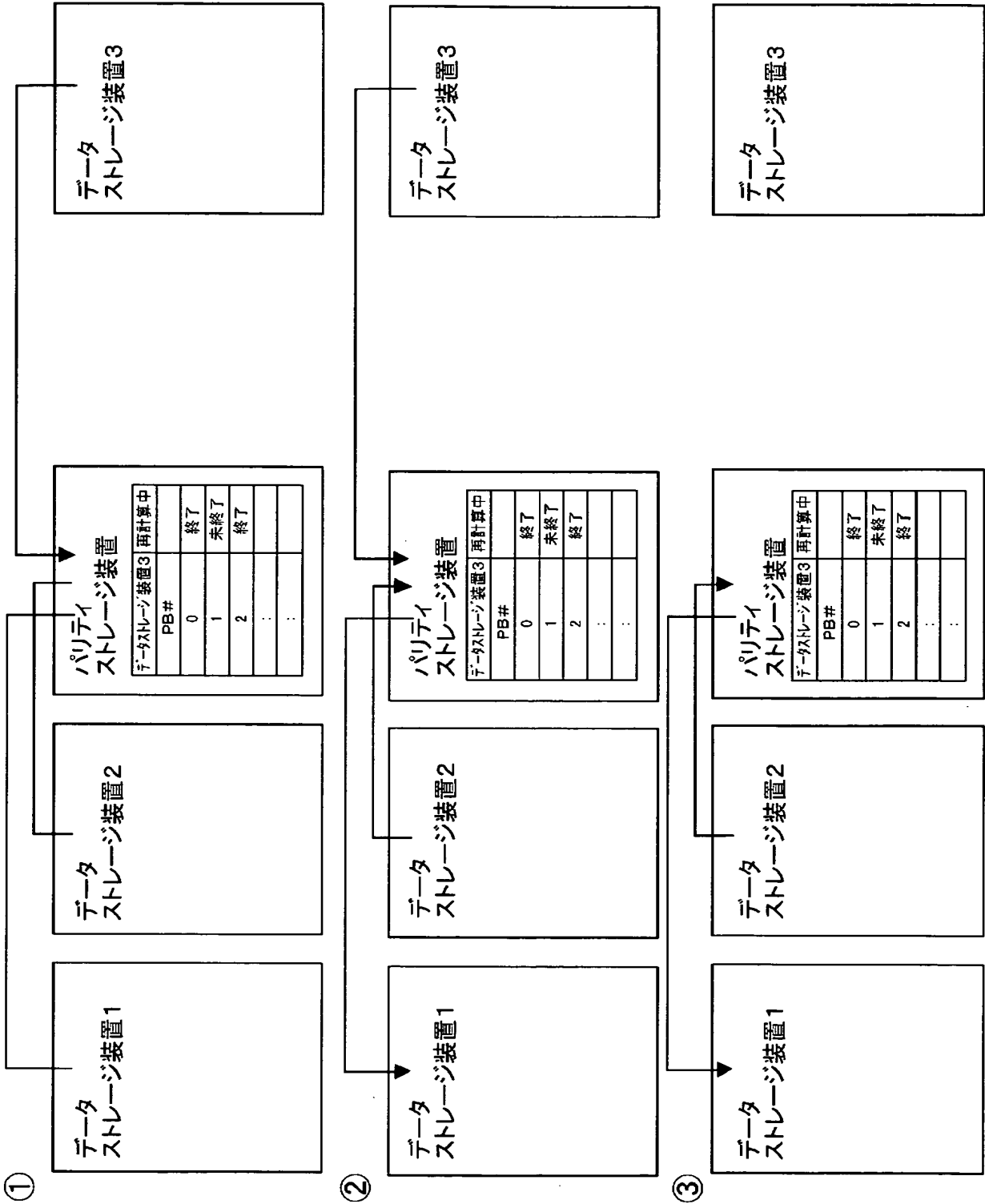




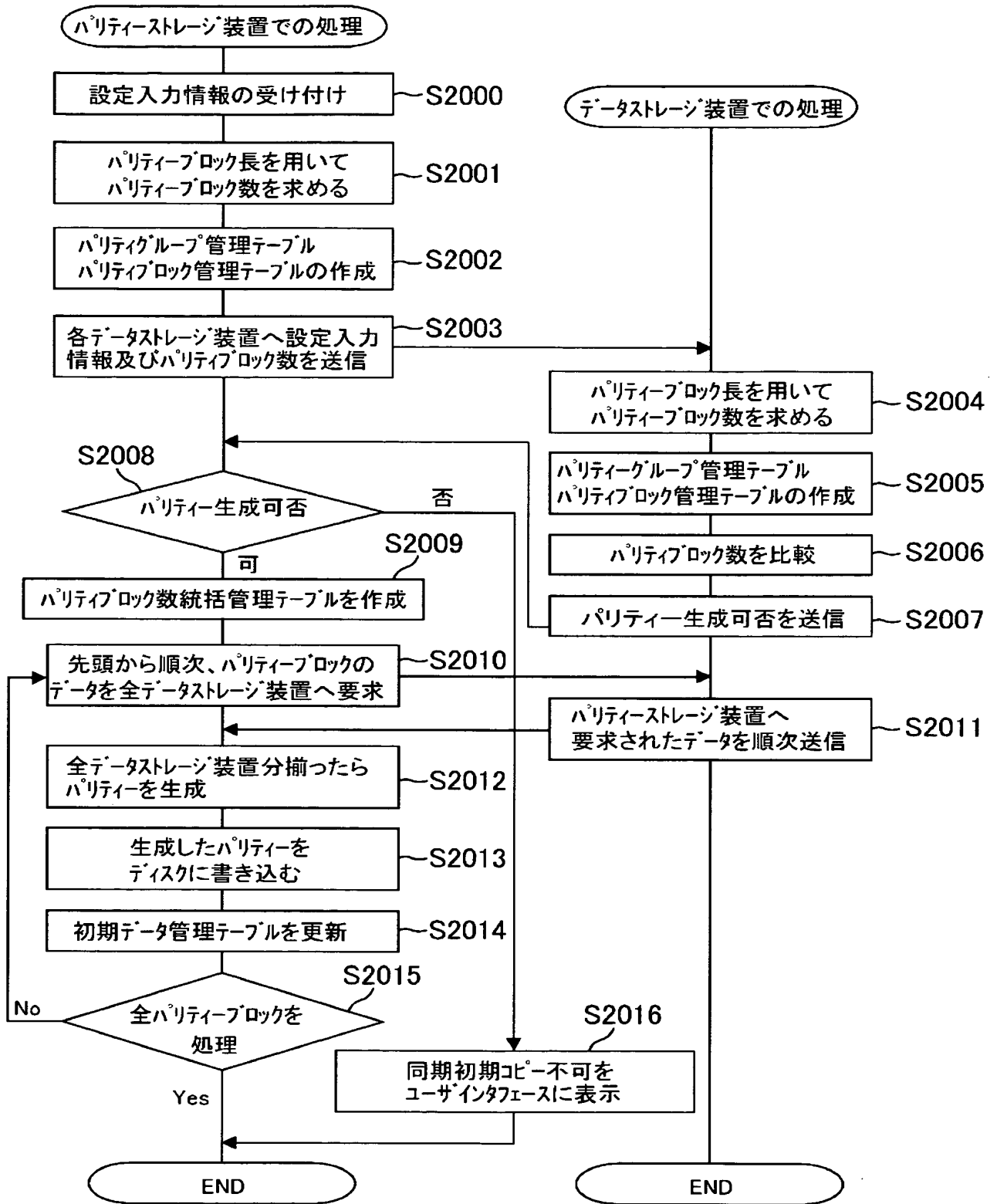
【図 19】



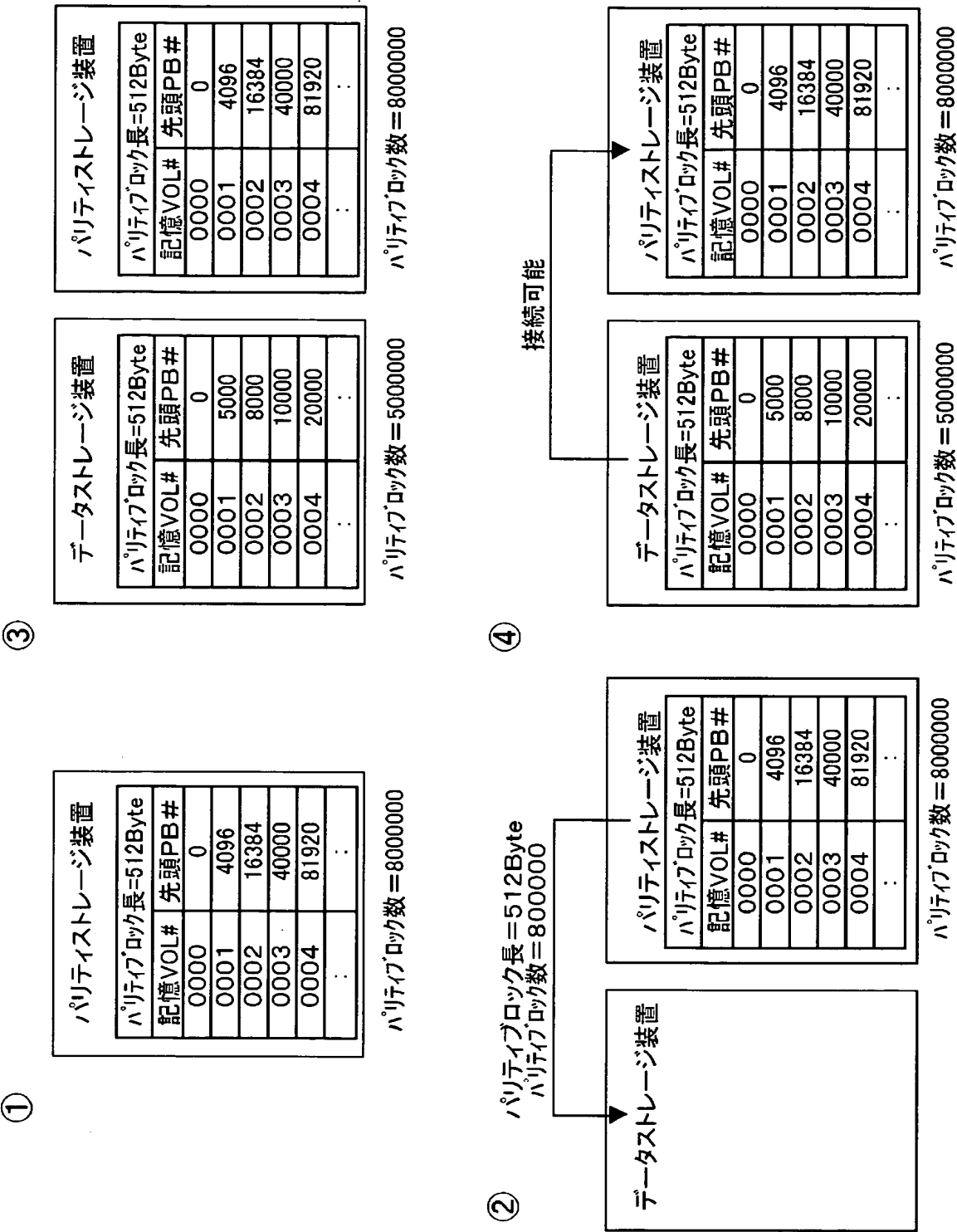
【図 20】



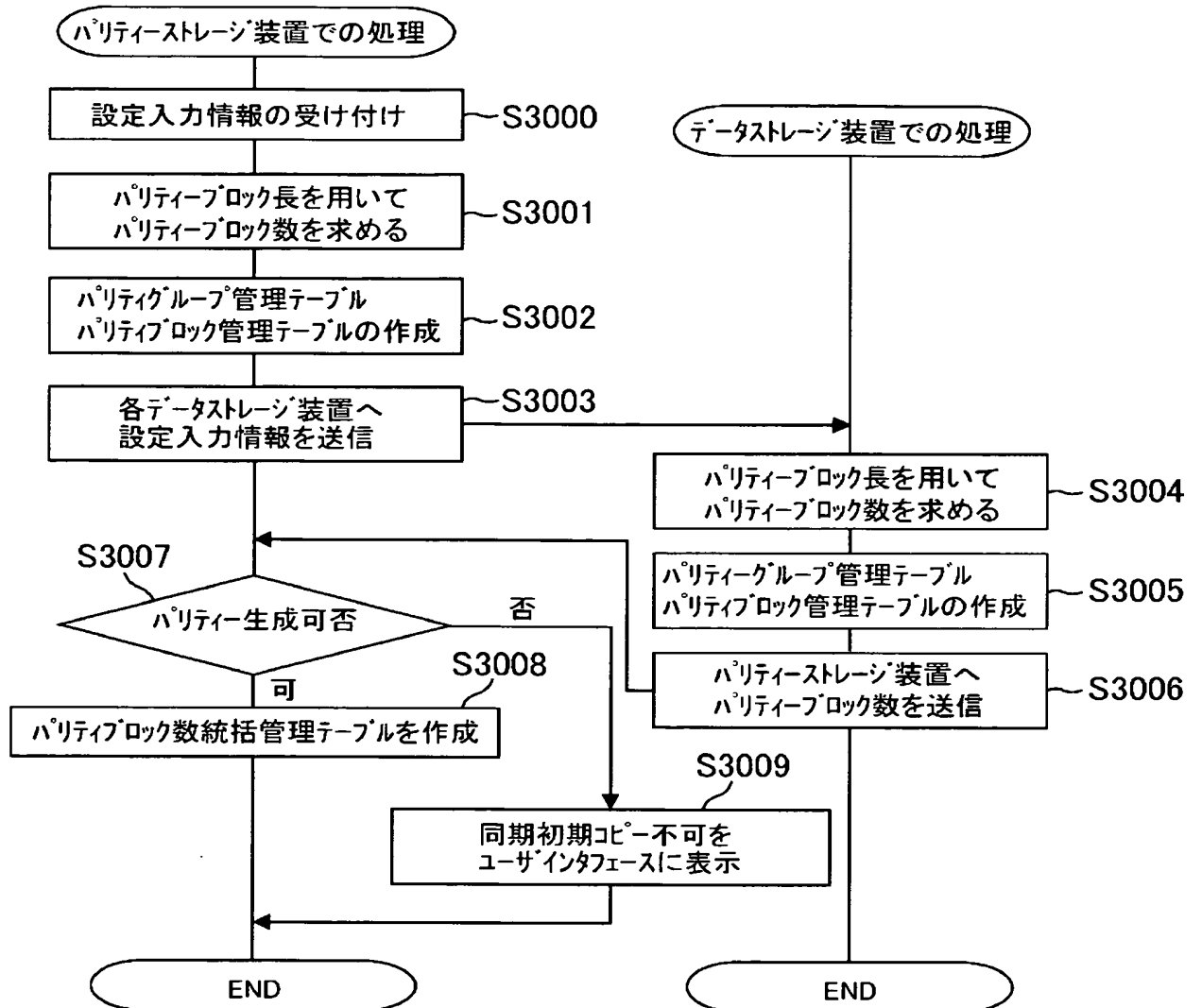
【図 21】



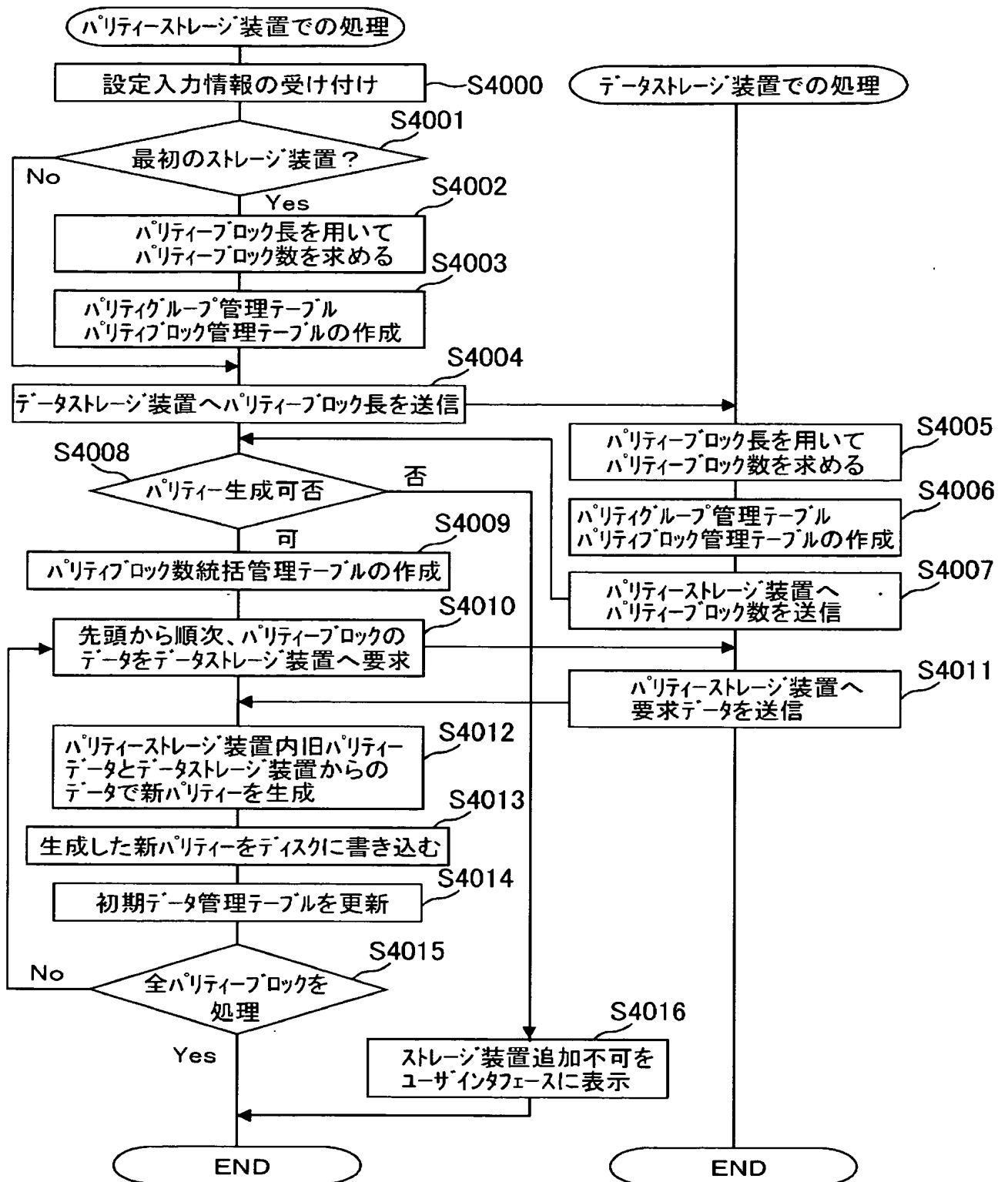
【図 22】



【図 23】



【図 24】



【図 25】

非同期方式パリティグループ設定画面

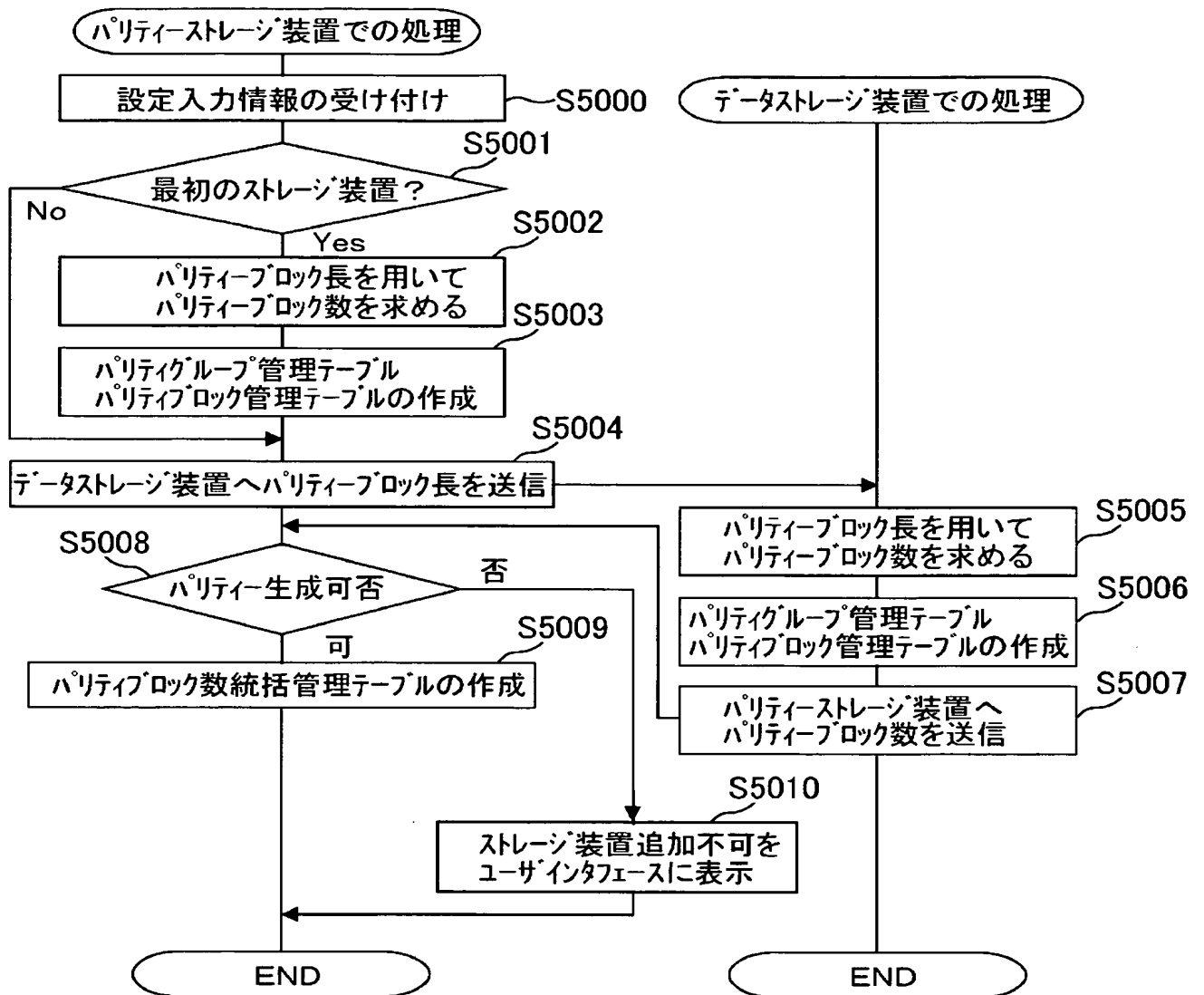
現在の構成	
パリティストレージ装置	ストレージ装置4
データストレージ装置	ストレージ装置1,ストレージ装置2
パリティブロック長	512バイト

追加するデータストレージ装置

初期化方式

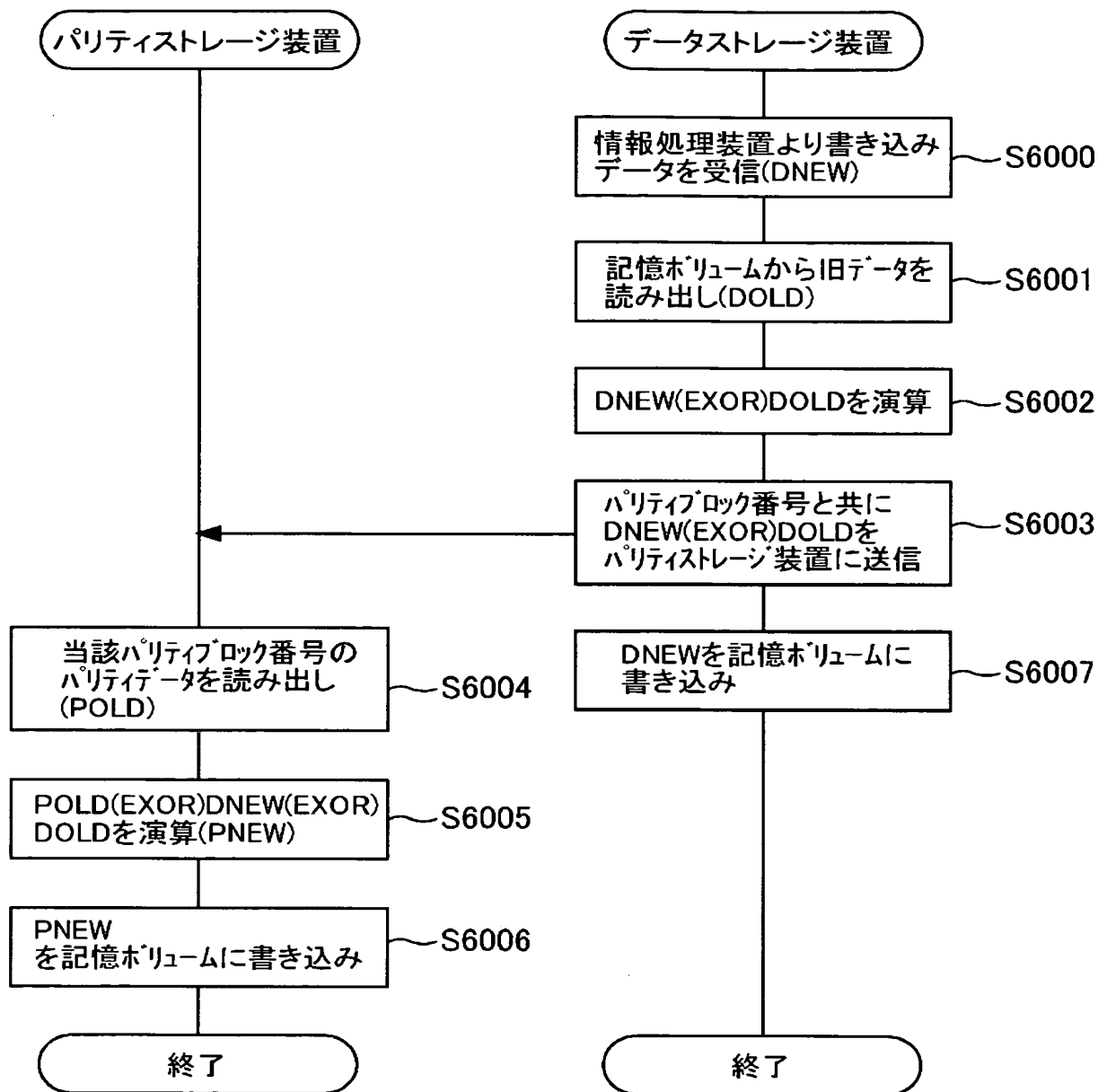
削除するデータストレージ装置

【図 26】





【図 27】

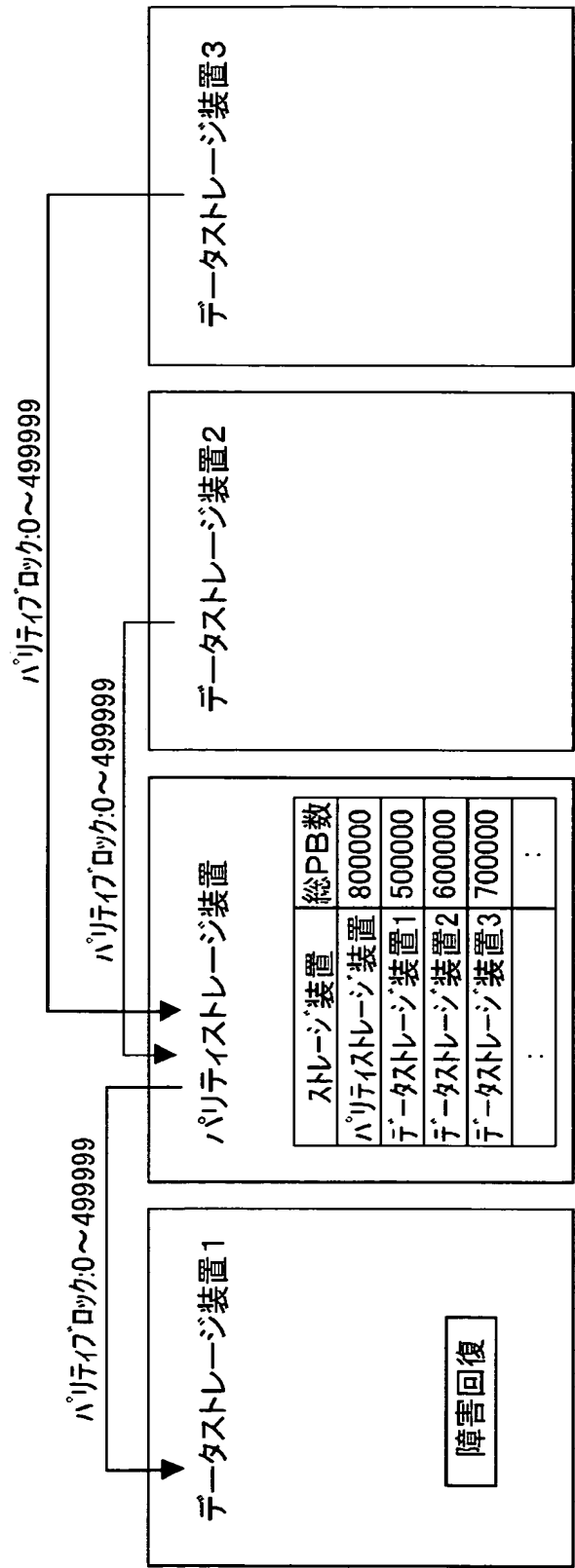


```

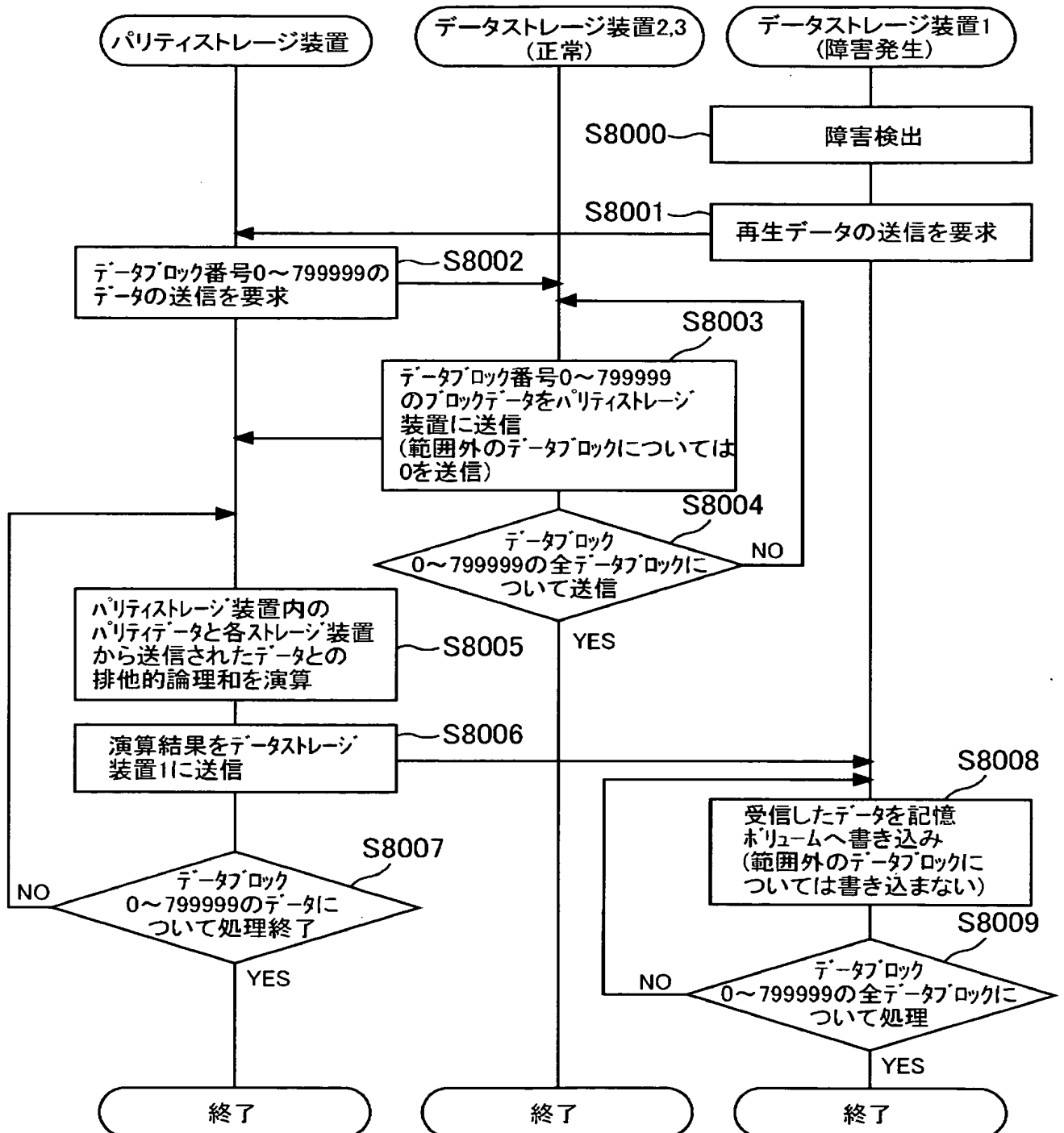
graph TD
    subgraph Devices
        P[パリティストレージ装置]
        D23[データストレージ装置2,3  
(正常)]
        D1[データストレージ装置1  
(障害発生)]
    end

    D1 -- S7000 障害検出 --> R[再生データの送信を要求]
    R -- S7001 --> P
    P -- S7002 --> D23
    D23 -- S7003 --> P
    P -- S7004 判定 --> YES1[YES]
    P -- S7004 判定 --> NO1[NO]
    YES1 -- S7005 --> A1[パリティストレージ装置内の  
パリティデータと各ストレージ装置  
から送信されたデータとの  
排他的論理和を演算]
    NO1 -- S7006 --> A2[演算結果をデータストレージ  
装置1に送信]
    A1 --> D1
    A2 --> D1
    A1 -- S7007 判定 --> YES2[YES]
    A1 -- S7007 判定 --> NO2[NO]
    YES2 --> End1([終了])
    NO2 --> P
    A2 -- S7008 --> B[受信したデータを記憶  
ボリュームへ書き込み]
    B -- S7009 判定 --> YES3[YES]
    B -- S7009 判定 --> NO3[NO]
    YES3 --> End2([終了])
    NO3 --> P
  
```

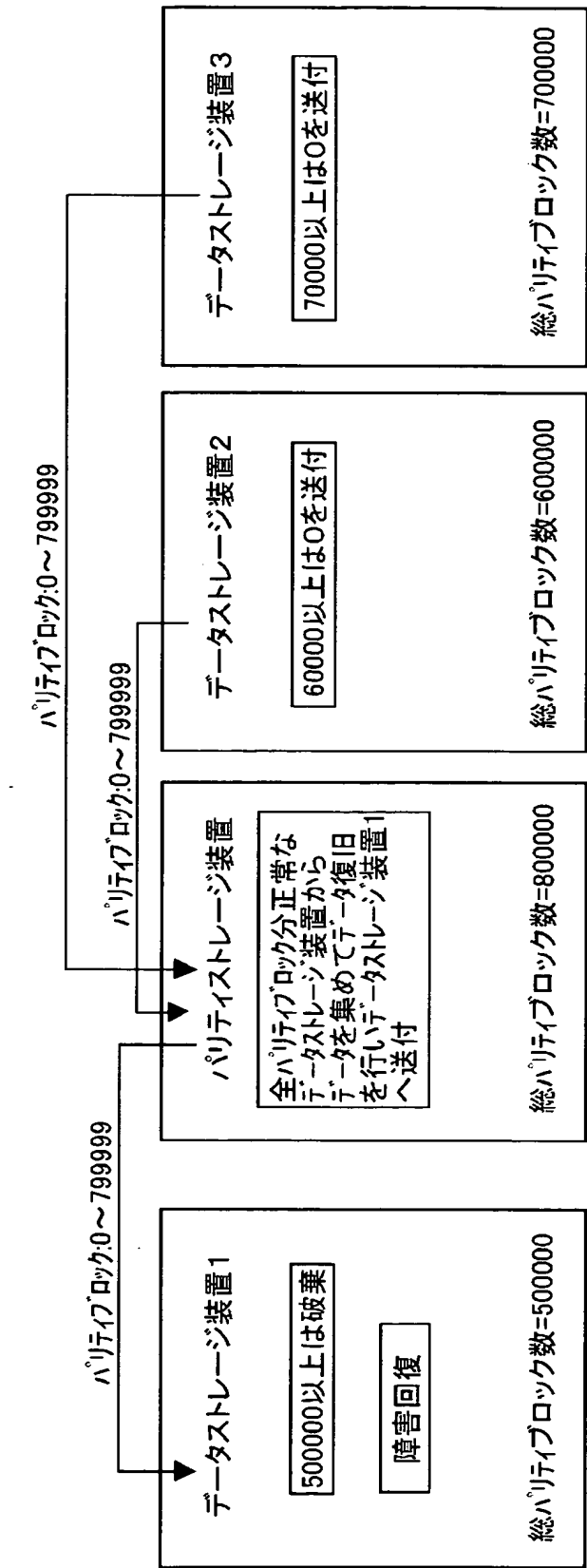
【図 29】



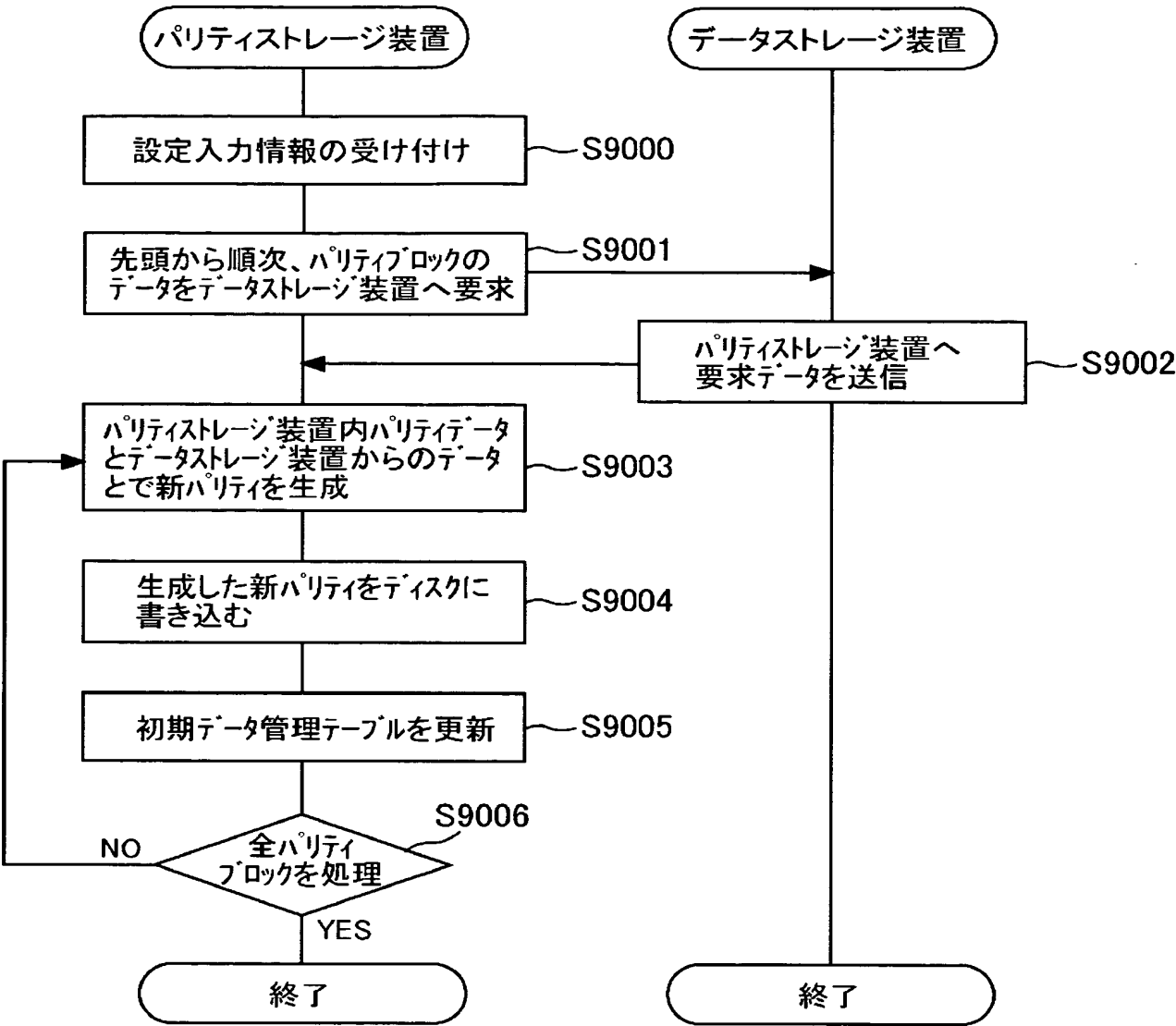
【図 30】



【図 31】



【図 3 2】



【図 3 3】

パリティブロック番号	ステータス情報	データ
------------	---------	-----

## 【書類名】 要約書

## 【要約】

【解決手段】 複数の第1ディスクドライブを有する複数の他のストレージ装置と通信可能に接続され、複数の第2ディスクドライブと、他のストレージ装置のそれぞれから複数の第1ディスクドライブの記憶領域を論理的に区分してなる複数の記憶ブロックのそれぞれに記憶される第1記憶データの複製を各記憶ブロックを特定する第1識別子と共に受信する第1受信部と、第1受信部により受信された第1記憶データの複製のうち第1識別子がそれぞれ対応する第1記憶データの複製の排他的論理和を演算する第1演算制御部と、複数の第2ディスクドライブの記憶領域を論理的に区分してなる複数の記憶ブロックのそれぞれを特定する第2識別子が第1識別子と対応する第2ディスクドライブの記憶ブロックに第1演算制御部により演算された排他的論理和の演算結果を記憶する第1記憶制御部とを備えるストレージ装置に関する。

【選択図】 図1

特願 2 0 0 4 - 0 1 9 7 3 9

出 願 人 履 歴 情 報

識別番号

[ 0 0 0 0 0 5 1 0 8 ]

1. 変更年月日

1 9 9 0 年 8 月 3 1 日

[変更理由]

新規登録

住 所

東京都千代田区神田駿河台 4 丁目 6 番地

氏 名

株式会社日立製作所